

# Determination of Complex Reaction Mechanisms

Analysis of Chemical, Biological,  
and Genetic Networks

---

John Ross

Igor Schreiber

Marcel O. Vlad

WITH CONTRIBUTIONS FROM

Adam Arkin

Peter J. Oefner and Nicola Zamboni

# Determination of Complex Reaction Mechanisms

*This page intentionally left blank*

# Determination of Complex Reaction Mechanisms

Analysis of Chemical, Biological,  
and Genetic Networks

---

John Ross

*Department of Chemistry, Stanford University*

Igor Schreiber

*Department of Chemical Engineering,  
Prague Institute of Chemical Technology*

Marcel O. Vlad

*Department of Chemistry, Stanford University, and  
Center for Mathematical Statistics, Romania*

WITH CONTRIBUTIONS FROM

Adam Arkin

*Department of Bioengineering,  
University of California, Berkeley*

Peter J. Oefner and Nicola Zamboni

*Stanford Genome Technology Center,  
Stanford University*

OXFORD

UNIVERSITY PRESS

2006

OXFORD  
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further  
Oxford University's objective of excellence  
in research, scholarship, and education.

Oxford New York  
Auckland Cape Town Dar es Salaam Hong Kong Karachi  
Kuala Lumpur Madrid Melbourne Mexico City Nairobi  
New Delhi Shanghai Taipei Toronto

With offices in  
Argentina Austria Brazil Chile Czech Republic France Greece  
Guatemala Hungary Italy Japan Poland Portugal Singapore  
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2006 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.  
198 Madison Avenue, New York, New York 10016  
www.oup.com

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,  
stored in a retrieval system, or transmitted, in any form or by any means,  
electronic, mechanical, photocopying, recording, or otherwise,  
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data

Ross, John, 1926–

Determination of complex reaction mechanisms : analysis of chemical, biological, and genetic networks / John Ross, Igor  
Schreiber, Marcel O. Vlad ; with contributions from Adam Arkin, Peter J. Oefner, and Nicola Zamboni.  
p. cm.

Includes bibliographical references and index.

ISBN-13: 978-0-19-517868-5

ISBN 0-19-517868-8

1. Chemical kinetics. 2. Reaction mechanisms (Chemistry) 3. Chemical reaction,  
Conditions and laws of. I. Schreiber, Igor. II. Vlad, Marcel O. III. Title.  
QD502.5.R67 2005

541'.394—dc22 2005001379

9 8 7 6 5 4 3 2 1

Printed in the United States of America  
on acid-free paper

*We dedicate this book to  
our students,  
our coworkers,  
and our families*

*This page intentionally left blank*

# Contents

---

- 1 Introduction, 1
  - 1.1 Some Basic Definitions, 3
  - 1.2 Introduction to Classical Identification, 4
- 2 Introduction to Chemical Kinetic Processes, 7
  - 2.1 Macroscopic, Deterministic Chemical Kinetics, 7
  - 2.2 Disordered Kinetics, 9
  - 2.3 Fluctuations, 9
- 3 A Brief Review of Methodology for the Analysis of Biochemical Reactions and Cells, 11
  - 3.1 Introduction, 11
  - 3.2 Measurement of Metabolite Concentrations, 11
  - 3.3 Principles and Applications of Mass Spectrometry, 15
  - 3.4 Genome-Wide Analyses of mRNA and Proteins, 18
  - 3.5 Fluorescent Imaging, 23
  - 3.6 Conclusions, 27
- 4 Computations by Means of Macroscopic Chemical Kinetics, 34
  - 4.1 Chemical Neurons and Logic Gates, 35



- 4.2 Implementation of Computers by Macroscopic Chemical Kinetics, 37
- 4.3 Computational Functions in Biochemical Reaction Systems, 40
- 5 Response of Systems to Pulse Perturbations, 46
  - 5.1 Theory, 46
  - 5.2 An Example: The Glycolytic Pathway, 53
- 6 Experimental Test of the Pulse Perturbation Method for Determining Causal Connectivities of Chemical Species in a Reaction Network, 58
- 7 Correlation Metric Construction: Theory of Statistical Construction of Reaction Mechanisms, 65
- 8 Experimental Test and Applications of Correlation Metric Construction, 87
- 9 Density Estimation, 97
  - 9.1 Entropy Metric Construction (EMC), 97
  - 9.2 Entropy Reduction Method (ERM), 102
- 10 Applications of Genetic Algorithms to the Determination of Reaction Mechanisms, 104
  - 10.1 A Short Primer on Genetic Algorithms, 104
  - 10.2 Selection of Regulation of Flux in a Metabolic Model, 106
  - 10.3 Evolutionary Development of Biochemical Oscillatory Reaction Mechanisms, 112
  - 10.4 Systematic Determination of Reaction Mechanism and Rate Coefficients, 119
  - 10.5 Summary, 122
- 11 Oscillatory Reactions, 125
  - 11.1 Introduction, 125
  - 11.2 Concepts and Theoretical Constructs, 127
  - 11.3 Experiments Leading to Information about the Oscillatory Reaction Mechanism, 138
  - 11.4 Examples of Deduction of Reaction Mechanism from Experiments, 151
  - 11.5 Limits of Stoichiometric Network Analysis, 166

- 12 Lifetime and Transit Time Distributions and Response Experiments in Chemical Kinetics, 170
  - 12.1 Lifetime Distributions of Chemical Species, 171
  - 12.2 Response Experiments and Lifetime Distributions, 173
  - 12.3 Transit Time Distributions in Complex Chemical Systems, 182
  - 12.4 Transit Time Distributions, Linear Response, and Extracting Kinetic Information from Experimental Data, 188
  - 12.5 Errors in Response Experiments, 192
  - 12.6 Response Experiments for Reaction–Diffusion Systems, 197
  - 12.7 Conclusions, 205
- 13 Mini-Introduction to Bioinformatics, 207
  - 13.1 Clustering, 209
  - 13.2 Linearization in Various Forms, 209
  - 13.3 Modeling of Reaction Mechanisms, 210
  - 13.4 Boolean Networks, 211
  - 13.5 Correlation Metric Construction for Genetic Networks, 213
  - 13.6 Bayesian Networks, 215
  - 13.7 Some Other Illustrative Approaches, 220
- Index, 223

*This page intentionally left blank*

# Introduction

---

Chemical kinetics as a science has existed for more than a century. It deals with the rates of reactions and the details of how a given reaction proceeds from reactants to products. In a chemical system with many chemical species, there are several questions to be asked: What species react with what other species? In what temporal order? With what catalysts? And with what results? The answers constitute the macroscopic reaction mechanism. The process can be described macroscopically by listing the reactants, intermediates, products, and all the elementary reactions and catalysts in the reaction system.

The present book is a treatise and text on the determination of complex reaction mechanisms in chemistry and in chemical reaction systems that occur in chemical engineering, biochemistry, biology, biotechnology, and genomics. A basic knowledge of chemical kinetics is assumed.

Several approaches are suggested for the deduction of information on the causal chemical connectivity of the species, on the elementary reactions among the species, and on the sequence of the elementary reactions that constitute the reaction pathway and the reaction mechanism. Chemical reactions occur by the collisions of molecules, and such an event is called an elementary reaction for specified reactant and product molecules. A balanced stoichiometric equation for an elementary reaction yields the number of each type of molecule according to conservation of atoms, mass, and charge. Figure 1.1 shows a relatively simple reaction mechanism for the decomposition of ozone by light, postulated to occur in a series of three elementary steps. (The details of collisions of molecules and bond rearrangements are not discussed.) All approaches are based on the measurements of the concentrations of chemical species in the whole reaction system, not on parts, as has been the practice.

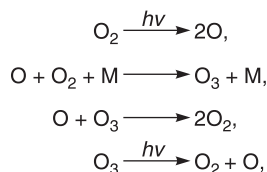


Fig. 1.1 Hypothesis for the elementary steps in the formation and decomposition reactions of ozone irradiated by light in the upper atmosphere.

One approach is called the pulse method, in which a pulse of concentration of one or more species of arbitrary strength is applied to a reacting system and the responses of as many species as possible are measured. From these responses causal chemical connectivities may be inferred. The basic theory is explained, demonstrated on a model mechanism, and tested in an experiment on a part of glycolysis.

In a second approach, measurements on time series of concentrations are used to construct correlation functions of concentrations. A theory is developed that shows that from these correlation functions information may be inferred on the reaction pathway, the reaction mechanism, and the centers of control in that mechanism. This method has been tested in experiments on glycolysis and has been applied to an extensive experimental genome study.

A third approach is based on the application of genetic algorithm methods to the study of the evolutionary development of a reaction mechanism, to the attainment of given goals (flux control, for example) in a mechanism, and to the determination of a reaction mechanism and rate coefficients by comparison with experiments.

Oscillatory reactions require a separate analysis, which is presented in detail. Responses of nonlinear systems to pulses or other perturbations are treated in some generality. The concluding chapter gives a brief introduction to bioinformatics, including several methods for determining reaction mechanisms.

There is not much literature (publications, texts, or treatises) on discussion of the methods and the art of guessing reaction mechanisms. In [1] there is a listing of hints and several rules that are helpful for developing a reaction mechanism for relatively simple systems. Once you have guessed a series of elementary steps, then there are two approximate methods that have been used extensively. One method is applicable for systems with a number of elementary reactions, one of which is slower than the others and hence is said to be rate-determining. The other is applicable for systems in which all the elementary reactions have about the same rate, but the time variation of intermediates is close to zero, the so-called stationary state hypothesis [1]. These two approximations have been very useful.

For more complex systems the approach used for many years, in its simplest description, has been the following: Identify as many species in the system as possible; separate the system into individual elementary steps; measure the stoichiometric coefficient for each species in each elementary step; determine the kinetics (the rate) of each elementary step; and postulate how the elementary steps may follow each other in sequence to produce the final products from the initial reactants. From this postulated reaction mechanism, you may derive kinetic equations and check these against available experiments. If there is agreement, then the mechanism is sufficient—it is a possible mechanism. In addition to this simple outline many other techniques have been used with great success: isotope measurements, mass spectrometry, and ion cyclotron methods, among many others (see chapter 3).

The approaches and methods discussed here differ from the prior art described in the last paragraph. The attempt is to keep as much of the system together as possible in order to maintain as many of the connections as possible: the connections of the species to each other, the connections of the effectors on enzymes, and so on. The isolation of a single elementary step cuts many interactions, as seen in fig. 1.2: isolate the reaction fructose 6-phosphate to fructose 1,6-bisphosphate, and the interactions with fructose 2,6-bisphosphate, among many others, are severed. “To keep as much of the system together as possible” means to measure as many species as possible. The enormous progress in the instrumentation for the measurements of chemical species and their concentrations in the last twenty-five years now permits the simultaneous measurements of dozens to hundreds of species (see chapter 3); we need the construction of experiments and theories designed to deduce or infer, not guess, from such measurements information on the connectivity of the species, on the reaction pathway, and on the reaction mechanism.

As chemistry penetrates biology, biotechnology, and genome studies, the reaction systems being considered become more and more complex, with many more species, many more catalysts, more complex connectivities, and positive and negative feedforward and feedback interactions. New experiments and theories are necessary for investigating such systems, and this book is an introduction to the field. Much, but certainly not all, of the contents of the book is the result of our research studies in the last fifteen years, and it is time to assemble this work into a coherent presentation.

## 1.1 Some Basic Definitions

We need to begin with several definitions of terms used throughout. (For an introduction to chemical kinetics see [1].)

*Elementary reaction.* The term “elementary reaction” has been defined. Sometimes two or more elementary reactions are added, as for example the elementary steps in a Michaelis–Menten enzymatic reaction, and listed as a pseudo-elementary reaction.

*Catalyst.* A catalyst may be necessary for an elementary reaction to occur at a reasonable rate, as well as positive and negative effectors on the catalyst. We shall be concerned mostly with isothermal reactions.

*Reaction mechanism.* A series of elementary reaction steps, usually postulated, that leads from stated reactants to products, including catalysts and effectors, is called the reaction mechanism for that overall reaction. An example of an important biological reaction mechanism, a part of glycolysis, is shown in fig. 1.2.

*Reaction pathway.* A listing of sequential chemical species in a reaction mechanism is called a reaction pathway; for example, the sequence glucose 6-phosphate, fructose 6-phosphate, fructose 1,6-bisphosphate (see fig. 1.2) is a reaction pathway in the glycolysis reaction mechanism.

*Causal connectivity.* The direct connection of one species, say glucose, to another, glucose 6-phosphate, by reaction is called causal connectivity. A sequence of causal connectivities gives a reaction pathway from which a reaction mechanism, or parts thereof, may be inferred. There will be more on this subject later.

*Control.* A mechanism may be designed for control (regulation) of various functions: control of concentrations (homeostasis), of rates, of specific product formation, of temperature, and so on. We shall see some examples later.

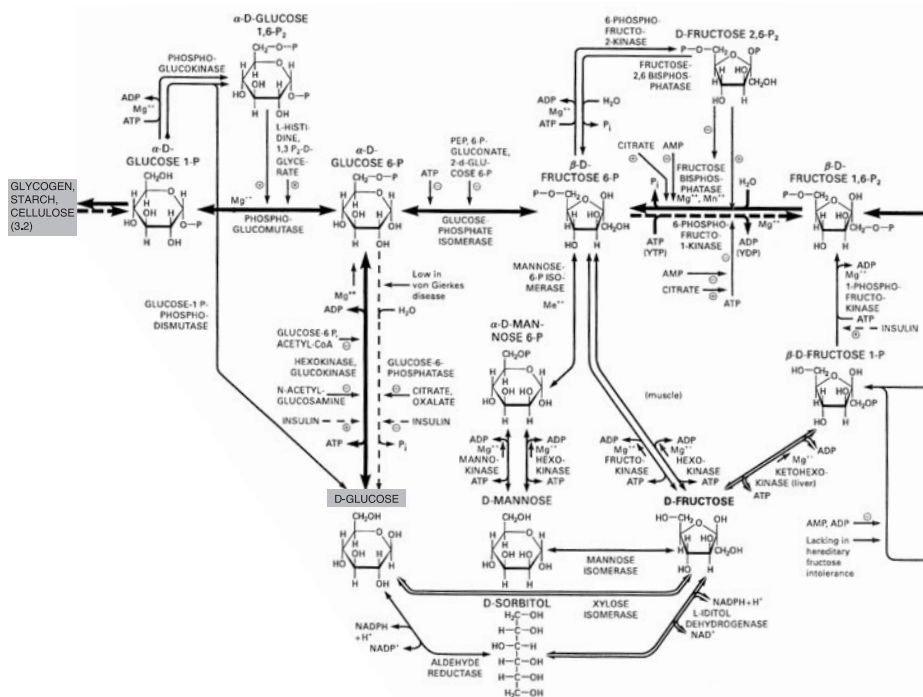


Fig. 1.2 Details of the reaction mechanisms of a part of glycolysis and the pentose cycle. Note the presence of the metabolites, enzymes, and the positive and negative effectors on the enzymes. Species will be defined in the later figures and in the text as needed. (Reproduced with permission from [8].)

Let us deal with the issue of uniqueness of a reaction mechanism. If a scientific theory, model, or hypothesis agrees with available experiments, then the agreement provides sufficiency, never necessity, for these theoretical constructs. In the literature of chemical kinetics there are ubiquitous warnings that the agreement of the predictions of a reaction mechanism with experiment does not prove that the mechanism is unique, that is, necessary, but only that it is a possible mechanism, it is sufficient. You can show that a mechanism is wrong by comparison with experiments; you cannot prove that it is necessary. These statements hold also for our approaches, as for all of science; they yield sufficient, not necessary, pathways and mechanisms.

## 1.2 Introduction to Classical Identification

The purpose of this section is to expose the reader to the acknowledged scholarly field of deducing, or inferring, structure and dynamics from measurements, also called classical identification (see [2] and [3] for representative references). This is well known to

chemical engineers, but strangely the topic cannot be found in the chemical literature and texts. Let us consider an example. Suppose you feed a batch culture of some bacteria with suitable nutrients at a given inflow rate, and you measure a series of molecular (concentrations of species) and physiological (pH, temperature) functions in time; you may also measure gene expression (protein production) and the production of some drug or secondary metabolite. Say that you wish to optimize the rate of production of the drug by varying the inflow of nutrients and varying concentrations of intermediates, pH, temperature, and so on. We can express the variations of the concentrations of the genes, the metabolites, and the drug by an equation in which general model structure is asserted. An example is a simple nonlinear regression in which the observation of one variation at time  $t$  is dependent on other variables at time  $(t - \tau)$ , and may be expressed as:

$$S_i = \beta_0 + \sum_j \beta_{1,j} S_j(t - \tau) + \sum_j \sum_k \beta_{2,j,k} S_j(t - \tau) S_k(t - \tau) + \cdots \quad (1.1)$$

where the regression methods solve for the parameters  $\beta$ .

This example is relatively unstructured and methods like LISREL (reference in the next paragraph) allow the assertion of more causal hypotheses into the regression. We may simply search for the values of the parameters  $\beta$  which maximize the production of the desired drug, say  $S_3$ , without seeking causal connections. We may also guess terms in this equation from prior knowledge or eliminate terms by designing gene deletion experiments and thus introduce causal relations. Clearly, as more and more causal structure is entered, sooner or later one arrives at a full mechanistic model. In all cases, the problem reduces to one of parameter estimation for the coefficients of the equations. In chapters 5, 6, and 12 we describe approaches based on perturbing a system, linear or nonlinear, measuring responses, and inferring information on the reaction mechanisms of the system.

Analysis by linear regression is the fit of given measurements by a straight line. The programs LISREL and SEPATH (see [4]–[7]) provide linear regression with some causal input. The linear or linearized response of a system to a small perturbation depends on the location of the system in the space of its variables. To obtain a linear response in a possibly nonlinear system requires small perturbations, and hence to be useful the errors in the measurements must be small—smaller than the perturbations. The issues of importance here are:

1. *Observability*. Perturb one species and see what you can determine, that is, measure, about the other species.
2. *Controllability*. For a given system, can you perturb at a point A in the space of the variables and from there come to another point B?
3. *Reachability*. Is the attainment of point B from point A feasible?

Kalman filtering is used for nonlinear systems either to control a given output or to fit a set of measurements by a linear fit over a small region of the measurements and vary that fit from one region to the next.

The classical identification becomes difficult or inapplicable when we have systems with a high degree of nonlinearity, when only partial observations can be made, and when there are stochastic (random) elements, that is, the variables are given by probability distributions rather than by given numbers. Multivariate systems in which only relatively inaccurate and imprecise measurements can be made, usually with insufficient time and



space resolution, may require probabilistic rather than deterministic interpretations and analyses. This type of statistical approach is presented in chapters 7 and 8 and in parts of chapter 13, where some elements of Bayesian statistics are introduced. The inference of causality may then be difficult or impossible, as many mechanisms may lead to the same joint probability distribution of all events in the system.

If measurements in time and space can be made at smaller and smaller intervals, then ultimately we come to molecular events, such as reactive collisions, and to a sequence of causal connectivities. To some this seems an obvious fact, to others it is a belief; the pursuit of these issues enters the field of philosophy.

A final point needs to be made regarding taking measurements in time. Suppose we sample a signal  $y(t)$  at a sampling rate of  $\Delta$ . Nyquist showed that if there exists a critical frequency  $f_c = 1/(2\Delta)$  for a system such that measurements are limited to frequencies smaller than  $f_c$ , then the function  $y(t)$  is completely determined by these measurements.

*Acknowledgments* Much of this book is based on our research, with appropriate references, and we copy freely from our publications with additions, deletions, and changes. The use of “we” refers to the respective authors of the referenced section. The contributions of our coauthors are gratefully acknowledged.

## References

- [1] Berry, R. S.; Rice, S. A.; Ross, J. *Physical Chemistry*, 2nd ed.; Oxford University Press: Oxford, 2000.
- [2] Ljung, L. *System Identification: Theory for the User*, 2nd ed.; Prentice-Hall: Englewood Cliffs, NJ, 1999.
- [3] Nelles, O. *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*; Springer: Berlin, 2001.
- [4] Broekstra, G. C-analysis of C-structures: representation and evaluation of reconstruction hypotheses by information measures. *Int. J. Gen. Sys.* **1981**, 7, 33–61.
- [5] Marriot, F. H. C. *The Interpretation of Multiple Observations*; Academic Press: San Francisco, 1979.
- [6] Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes in C*; Cambridge University Press: New York, 1988.
- [7] Otnes, R. K.; Enochson, L. *Applied Time Series Analysis: Basic Technique*; Wiley: New York, 1978.
- [8] Michael, G. *Biochemical Pathways*; Elsevier/Spektrum Akademischer Verlag: Heidelberg, 1998.

## 2

# Introduction to Chemical Kinetic Processes

---

It is useful to have a brief discussion of some kinetic processes that we shall treat in later chapters. Some, but not all, of the material in this chapter is presented in [1] in more detail.

## 2.1 Macroscopic, Deterministic Chemical Kinetics

A macroscopic, deterministic chemical reacting system consists of a number of different species, each with a given concentration (molecules or moles per unit volume). The word “macroscopic” implies that the concentrations are of the order of Avogadro’s number (about  $6.02 \times 10^{23}$ ) per liter. The concentrations are constant at a given instant, that is, thermal fluctuations away from the average concentration are negligibly small (more in section 2.3). The kinetics in many cases, but far from all, obeys mass action rate expressions of the type

$$\frac{dA}{dt} = k(T)A^\alpha B^\beta \dots \quad (2.1)$$

where  $T$  is temperature,  $A$  is the concentration of species  $A$ , the same for  $B$ , and possibly other species indicated by dots in the equation, and  $\alpha$  and  $\beta$  are empirically determined “orders” of reaction. The rate coefficient  $k$  is generally a function of temperature and frequently a function of  $T$  only. The dependence of  $k$  on  $T$  is given empirically by the Arrhenius equation

$$k(T) = C \exp^{-E_a/RT} \quad (2.2)$$

where  $C$ , the frequency factor, is either nearly constant or a weakly dependent function of temperature, and  $E_a$  is the activation energy.

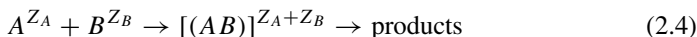
Rate coefficients are averages of reaction cross-sections, as measured for example by molecular beam experiments. The a priori calculation of cross-sections from quantum mechanical fundamentals is extraordinarily difficult and has been done to good accuracy only for the simplest trimolecular systems (such as  $D + H_2$ ).

A widely used alternative approach is based on activated complex theory. In its simplest form, two reactants collide and form an activated complex, said to be in equilibrium. One degree of freedom of the complex, a vibration, is allowed to lead to the dissociation of the complex to products, and the rate of that dissociation is taken to be the rate of the reaction. The rate for the forward reaction is

$$k_f = \frac{k_B T}{h} (K_V)' [A][B] \quad (2.3)$$

where  $k_B$  is Boltzmann's constant and  $K_V'$  is the equilibrium constant in terms of concentrations for the complex with one vibrational degree of freedom removed. On the right-hand side of eq. (2.3) there is frequently added a transmission coefficient to attempt to make up for the simplifications in that equation, due to (1) not all activated complexes dissociate to products, some dissociate back to reactants; (2) there may be quantum mechanical tunneling through the activation barrier separating reactants from products; and (3) the activated complex may not be at equilibrium and nonequilibrium effects may have to be taken into account.

Reactions among ionic species require explicit consideration of the interactions of the electric charges on the ions. We cite only the primary salt effect formulated by Brønsted and Bjerrum [1, p. 920]. For a bimolecular reaction between two ions we have



where  $Z_A$  and  $Z_B$  are the charges on the ions  $A$  and  $B$  respectively and  $(AB)$  is the activated complex. With the use of activated complex theory and the expression of the equilibrium constant for eq. (2.4) in terms of activities, we can write

$$K = \frac{a_{AB}}{a_A a_B} = \frac{[(AB)]}{[A][B]} \frac{\gamma_{AB}}{\gamma_A \gamma_B} \quad (2.5)$$

where the  $\gamma$ 's are activity coefficients, and the rate coefficient in the forward direction is

$$k_f = \frac{k_B T}{h} K \frac{\gamma_A \gamma_B}{\gamma_{AB}} \quad (2.6)$$

In dilute solutions the activity coefficient of ion  $i$  in water at 25°C is

$$\log \gamma_i = -0.509 Z_i^2 I^{1/2}, \quad I = \frac{1}{2} \sum_i Z_i^2 m_i \quad (2.7)$$

with  $m_i$  denoting molality.  $I$  is called the ionic strength. Hence the rate coefficient for the forward direction is

$$k_f = \frac{k_B T}{h} K \cdot 10^{1.018 Z_A Z_B I^{1/2}} \quad (2.8)$$

For reaction between two ions of like charge the rate coefficient increases with increasing ionic strength; for reaction between two ions of unlike charge the rate coefficient decreases with increasing ionic strength.

## 2.2 Disordered Kinetics

There are systems in which rate coefficients may vary randomly according to some probability distribution. Consider an amorphous solid in which an active intermediate is formed by radiation chemistry. The rate of reaction of that intermediate depends on the specific immediate surroundings of the intermediate, which may vary with location in the amorphous solid. The rate coefficients vary in space in the sample but not in time and such a system is said to have static disorder.

In systems with dynamic disorder, rate coefficients vary randomly in time; the structure of the environment of a reaction site varies as the reaction proceeds. Examples include protein–ligand interactions which vary in time as the protein changes its molecular configuration in time. Other examples are chemical processes in the atmosphere and aquatic environments. (For further analysis of disordered kinetics, see chapter 12.) A system may display both static and dynamic disorder depending on the external conditions. For example, in the case of protein–ligand interactions the reaction rates (rate coefficients) are random because a protein can exist in many different molecular configurations, each characterized by a different reaction rate. At low temperatures the transitions among different configurations are negligibly small (frozen out) and the system shows static disorder. For higher temperatures transitions among different configurations do occur, and the system shows dynamic disorder. For a few references on disordered kinetics, see [2]–[5].

## 2.3 Fluctuations

We make a distinction between internal, or thermal, fluctuations and external fluctuations, those due to variations in the environment of the system. Thermal fluctuations are present in every system at nonzero temperature. For systems at equilibrium the absolute magnitude of the fluctuations, say in concentration of a chemical species, increases with the number  $N$  of molecules of that species in a given volume, and is proportional to  $N^{1/2}$ . The relative fluctuations therefore decrease proportional to  $N^{-1/2}$ . At equilibrium the probability of a fluctuation of small magnitude is given by a Gaussian (normal) distribution. These relations lead to better understanding of the concept of a macroscopic system in which fluctuations are small compared to average quantities [1].

For nonequilibrium systems the probability of a fluctuation is given by the solution of the master equation. The absolute and relative fluctuations, at least for some nonequilibrium systems, are of similar magnitude to equilibrium systems.

As the number of molecules in a system at a given volume increases, the averages of the concentrations, averaged over the probability distribution, begin to obey mass action kinetics. For systems with mass action kinetics, whether at equilibrium or not, relative fluctuations tend to zero in the thermodynamic limit (the limit of large  $N$  at constant density  $N/V$ ), and are said to be nonintermittent. Rate processes in disordered systems have qualitatively different behavior. There is no universal behavior in the thermodynamic limit, but relative fluctuations tend to a constant nonzero value in that limit. These fluctuations are said to be intermittent and they make a significant contribution to the average values of the concentrations. For further references, see [6,7].

## References

- [1] Berry, R. S.; Rice, S. A.; Ross, J. *Physical Chemistry*, 2nd ed.; Oxford University Press: Oxford, 2000.
- [2] Shlesinger, M. Fractal time in condensed matter. *Annu. Rev. Phys. Chem.* **1988**, *39*, 269–290.
- [3] Plonka, A. *Time-Dependent Reactivity of Species in Condensed Media*; Springer: Berlin, 1986.
- [4] Zwanzig, R. Rate processes with dynamic disorder. *Acc. Chem. Res.* **1990**, *23*, 148–152.
- [5] Ross, J.; Vlad, M. O. Nonlinear kinetics and new approaches to complex reaction mechanisms. *Annu. Rev. Phys. Chem.* **1999**, *50*, 51–78. See the section on “Rate processes in disordered systems” and references therein.
- [6] Gardiner, C. W. *Handbook of Stochastic Methods*, 2nd ed.; Springer: New York, 1990.
- [7] Van Kampen, N. G. *Stochastic Processes in Physics and Chemistry*, 2nd ed.; Elsevier: Amsterdam, 1992.

# A Brief Review of Methodology for the Analysis of Biochemical Reactions and Cells

---

NICOLA ZAMBONI AND PETER J. OEFNER

## 3.1 Introduction

All cellular activities and responses to environmental stimuli are determined by a complex interplay of genes, RNA, proteins, and metabolites. As the observed phenotype is the result of all reactions occurring in the cell and with the environment, a comprehensive analytical representation of the biological system is required. The ideal analytical platform should identify and quantify, over time and in space, all possible species in a live cell and those trafficking with the environment. In addition, we expect the analysis not to interfere with the target. Because of the heterogeneity and number of species participating, such a platform does not exist yet. Nevertheless, the last decade has witnessed considerable technological advances, and a few systems have progressed sufficiently to support the understanding of cellular dynamics at a systemic level.

The goal of this review is to provide the reader with a short survey on the instrumentation and techniques that are mostly contributing to this cause, and to elucidate their principles and applications.

## 3.2 Measurement of Metabolite Concentrations

The experimental deduction of reaction mechanisms relies on the precise and accurate measurement of concentration changes of reacting species. With increasing system complexity, and thus number of analytes, separation methods are preferred to classical enzymatic assays, primarily for their capacity to analyze simultaneously a few to

hundreds of compounds in one experiment. When considering a whole cell, this task is very challenging because thousands of different species, the *metabolome*, coexist, with very different chemical and physical characteristics. Generally, targeted methods are designed to quantify specific species within a class of related intermediates, such as sugar phosphates, hexosamines, purines, amino acids, and lipids. Since targeted methods are tailored to the specific chemical or spectral properties of a given class, they lead to improved resolution and detection limits.

To date, most studies of cellular dynamics have focused on subsystems with a limited number of reactions or linked pathways. Analytes of interest typically belong to the same class and, thus, are accessible by a single analytical system. A prominent example are the intermediates of the central carbon metabolism, which comprises glycolysis, the pentose phosphate pathway, and the Krebs cycle. All metabolites in these pathways are carbohydrates that carry phosphate or carboxylic groups. Because of its pivotal role in catabolism and anabolism, in lower and higher organisms, central carbon metabolism is the best-studied biochemical system. Its analysis usually includes nucleotide phosphates, such as ATP and GTP, that reflect the energy state of a cell, and the redox cofactors  $\text{NAD}^+$  and  $\text{NADP}^+$ . These cofactors participate in a large number of reactions and interconnect metabolic pathways, acting as global regulators.

The analytical system of choice is expected to provide simultaneous access to the metabolites of interest and globally relevant cofactors. In addition, it is desirable to have a platform that can be flexibly extended for purposes of untargeted metabolome profiling. From this perspective, two conceptually different systems hold prime attention: high-pressure liquid chromatography and capillary electrophoresis.

### 3.2.1 High-Pressure Liquid Chromatography (HPLC)

Chromatography separates species dispersed in a mobile phase (gaseous or liquid) as a result of interactions with a static phase. The interactions can be of various kinds, for example, ionic or hydrophobic. Effective separations are obtained as a result of differences in retention on the stationary phase. Sugar phosphates (in glycolysis and the pentose phosphate pathway), carboxylic acids (in the Krebs cycle), and nucleotides are all negatively charged. This feature favors their simultaneous analysis by high-performance anion-exchange chromatography (HPAEC) [1–3]. In less than 1 h, these compound classes can be separated in complex samples such as those isolated from tissues, plant leaves, or microorganisms. Notably, in anion-exchange chromatography, the majority of substances with identical molecular weight in central metabolism have been successfully resolved. A possible drawback are the strong acidic or alkaline mobile phases, which can cause degradation of the labile phosphodiester bond in sugar phosphates [3] but are required to ensure the presence of all analytes in their ionic form.

Even more difficult than the separation is the detection of central carbon metabolites. Direct UV detection is suitable only for nucleotides and some Krebs cycle intermediates [1,2]. In contrast, amperometric detectors are efficient only for species bearing multiple hydroxylic groups such as sugars [3,4]. Refractometric, light scattering, or fluorescence and UV detection are valuable alternatives, but suffer from scarce versatility. The most comprehensive results have been obtained by conductometric detection, with detection limits in the low picomole range for both sugar phosphates and carboxylic acids [2,3].

This method allows the quantification of inorganic anions as well, although particular care is necessary to limit strong background signals originating from salt contaminations.

A breakthrough was achieved with the advent of electrospray interfaces for the coupling of HPLC to mass spectrometers (MS). We discuss MS in detail in a following section. In conjunction with chromatography, MS provides subpicomole sensitivity and high-resolution  $m/z$ -based separation of chromatographically unresolved analytes [5,6]. Historically, MS interfaces were first developed for gas chromatography (GC). GC-MS has in fact become a very robust system for the simultaneous analysis of several hundreds of compounds [7,8], and has been extensively used for profiling studies in biological extracts [9,10]. GC-MS has three major disadvantages: It degrades thermolabile compounds; it often requires derivatization steps because about 80% of endogenous metabolites are not volatile; and it is prone to in-source fragmentation.

Since these restrictions do not apply to HPLC-MS systems, efforts have been undertaken to extend the range of metabolites amenable to HPLC in response to the demand for alternatives to GC-MS. The synergy between HPAEC and MS has been successfully applied to the quantification of glycolytic intermediates [6]. However, ion-exchange chromatography involves high salt concentrations which generally interfere with the MS detection. To complement GC-MS, additional chromatographic modes have been adopted. Reverse-phase (RP) HPLC is perhaps the most important, and is widely used in the analysis of nonpolar compounds. In RP-HPLC, separation relies on hydrophobic interactions with aliphatic chains on the stationary phase in a mixed organic-aqueous mobile phase. RP-HPLC, with or without an ion-pairing reagent, was successfully employed to quantify the sugar phosphates in pentose phosphate pathway intermediates [5], and to profile plant metabolites [11,12]. On the other hand, MS-based analysis of highly polar analytes, including central carbon intermediates, peptides, oligosaccharides, or nucleic acids, can be achieved by hydrophilic-interaction chromatography (HILIC) [13].

### 3.2.2 Capillary Electrophoresis (CE)

CE refers to a group of related techniques that employ high DC voltages to drive within minutes the separation of the analytes in narrow-bore capillaries filled with weakly conductive buffers. In the classical configuration, a fused silica capillary is spanned between two buffer reservoirs, in each of which an electrode is immersed to apply an electric potential. In the presence of an electric field across the length of the capillary (up to 500 V/cm for 5–200  $\mu\text{m}$  i.d. capillaries), separation occurs by two overlapping phenomena: electrophoresis and electro-osmotic flow (EOF). In electrophoresis, ions migrate toward the electrode with opposite charge dependent on their mass, shape, and extent of solvation. In addition, all analytes, including neutral species, are affected by the EOF. The underlying mechanism is well understood: (1) The ionizable silanol groups (pI about 1.5) on the silica inner surface become negatively charged in contact with the background electrolyte (BGE) (at  $\text{pH} > \text{pI}$ ). (2) Cations from the BGE are then attracted to the walls, forming a double electrical layer. When a potential is applied at the ends of the capillary, the movement of the cations in the direction of the anode generates a flow of buffer solution. The strength of the EOF is a function of factors such as the ionic buffer strength, the density of charges on the capillary wall, the pH, the applied electric field, and the dielectric constant and the viscosity of the running buffer. In fact, the EOF



can be controlled for the specific requirements of the different CE modes by modifying these parameters or by chemical modification of the capillary inner surface. Although the effects of electrophoretic migration and EOF overlap, at neutral to alkaline pH the latter is overwhelming (up to 2–3 mm/s in silica capillaries) and ensures that cations, uncharged molecules, and anions migrate toward the anode at different velocities.

Detection typically occurs in-line, through an optical window placed in the proximity of the capillary outlet, either by spectral analysis, with average concentration sensitivities lower than in HPLC because of the shorter light path, or by laser-induced fluorescence [14], with detection limits reaching zeptomoles for appropriate fluorophores. Coupling and applications of CE-MS systems have been thoroughly reviewed [15]. For the analysis of complex biological mixtures, this combination is capable of detecting analytes such as proteins or metabolites in the femto- to attomole range.

CE offers important advantages compared to liquid chromatography [16]. In contrast to the parabolic flow profiles characteristic of hydrodynamic flow, such as in pumped HPLC, the EOF generated in electrophoresis is uniformly distributed across the tubing section and forms a plug-shaped flow profile. This drastically reduces band-broadening in CE, and results in very high resolution and, consequently, increased peak capacity. The major limitations are set by molecular diffusion and the heat generated by the resistance of the buffer to the current [15]. In addition, low sample volumes in the range of a few nanoliters are required. In the extreme case, CE was used to analyze the content of a single cell [17], corresponding to a volume of picoliters. Micromanipulation is typically used to introduce by suction a single cell in a 10–20  $\mu\text{m}$  i.d. capillary, in which cell lysis is induced by the hypotonic BGE and electrophoresis occurs [18]. Detection in the attomole range is required to measure species represented in the cells at concentrations of at least 1 mM, such as proteins [19,20], amino acids [21], and glycolytic endproducts [22].

In the family of CE, capillary zone electrophoresis (CZE) is the simplest and the most frequently used form of CE. The system is characterized by a homogeneous buffer solution and a constant field strength. CZE is routinely employed in the analysis of clinically relevant metabolites and drugs, inorganic ions, metals, and proteins [16,23,24]. Notably, Soga and coworkers combined three CZE-MS methods to isolate 1,692 compounds in bacterial extracts, of which 352 were identified and quantified [25] (fig. 3.1). Another form of CE is capillary isoelectric focusing (CIEF), in which the inner capillary wall is usually coated with polyacrylamide or methylcellulose to suppress the EOF. A series of zwitterionic species (the ampholytes) are used to generate a pH gradient in the capillary. In the absence of EOF, analytes and ampholytes migrate as long as they are charged. Consequently, they will separate and distribute along the gradient based on their isoelectric point (pI), where the net charge becomes zero and migration stops. In CIEF, high-resolution separations of proteins with pI differences < 0.01 units can be achieved. This renders CIEF particularly useful for monitoring variants of human and recombinant proteins or their glycosylation patterns. Micellar electrokinetic capillary chromatography (MECC) represents a unique chromatographic technique. In the BGE, surfactants (anionic, cationic, nonionic, or zwitterionic) are added in excess of their critical micellar concentration to form charged micelles. Analytes (also neutral) are separated based on their partitioning, and in the presence of chiral additives such as cyclodextrins, the separation is amenable to enantiomeric recognition [26]. Finally, affinity CE is frequently used in the characterization of biomolecular interactions between proteins, DNA, and

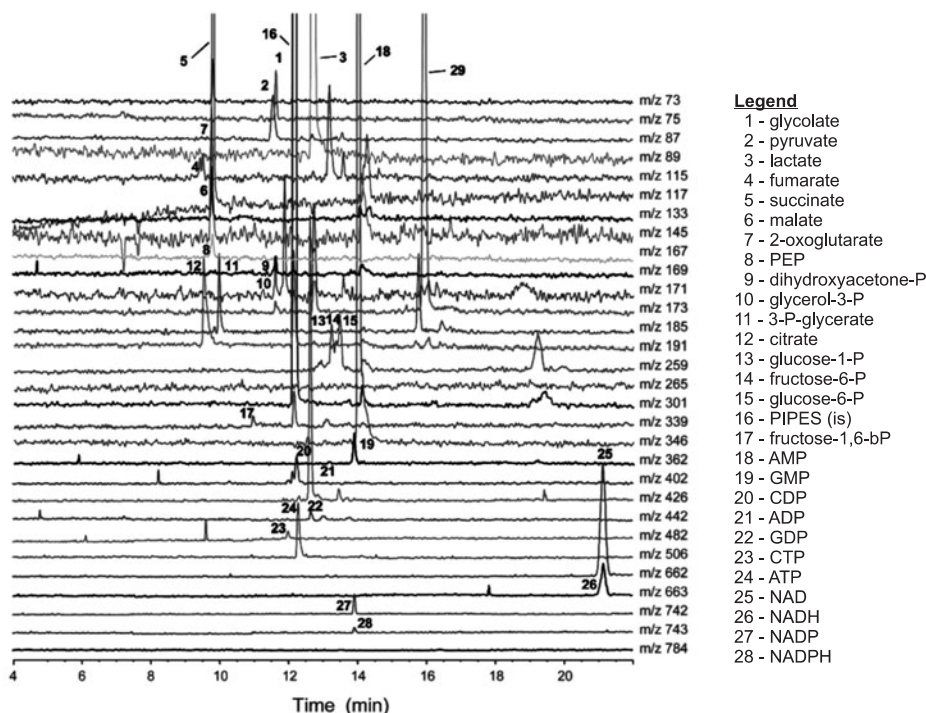


Fig. 3.1 CE-MS analysis of *Bacillus subtilis* extracts. Ion electropherograms are shown for selected intermediates of central carbon metabolism, energy equivalents, and redox cofactors. Separation was achieved in the negative mode by CZE in coated capillaries with reversed inner-wall polarity. (Reprinted with permission from [127].)

small molecules. In fact, binding constants, and in some cases stoichiometry and binding rates, are made measurable by affinity CE [27].

### 3.3 Principles and Applications of Mass Spectrometry

Mass spectrometry (MS) has many strengths: It is fast, robust, exquisitely sensitive, and offers high resolution; it can deliver structural information; and it is suited for the analysis of small molecules as well as large biopolymers [28]. In practical terms, the two prerequisites for MS measurement are that the analyte of interest is ionizable and introduced in the gas phase, because detection relies on acceleration and trapping of ions within electric and magnetic fields. Until the end of the 1980s the spectrum of natural species amenable to MS was severely restricted by their resistance to ionization, their nonvolatile nature, or the frequent fragmentation and thermal degradation occurring with strong ionization techniques. This changed with the introduction of two novel ionization interfaces: electrospray [29] and matrix-assisted laser desorption ionization [30,31].

These two techniques are characterized by high ionization efficiency for a wide range of biological molecules while preserving the analyte and enabling quantification.

Electrospray ionization (ESI) enables the transfer of ions from solution to the gas phase [32]. At the tip of the sprayer a strong electric field is imposed that causes formation of small charged droplets. Desolvation is obtained gradually by evaporation at ambient temperature and atmospheric pressure. Shrinking droplets are progressively destabilized by the increasing surface charge, and split into smaller droplets and ultimately gas-phase ions. ESI is widely used because of its compatibility with separation methods such as liquid chromatography and capillary electrophoresis. As a matter of fact, because of its flexibility with respect to the analyte and softness of ionization, ESI constitutes the ideal interface in on-line analyses of complex biological samples with LC-MS or CE-MS systems [32].

In matrix-assisted laser desorption/ionization (MALDI) the analytes are spotted on a surface in a thin layer of a matrix, typically an organic acid. A pulsed laser beam is used to form matrix vapor that lifts analyte molecules from the surface and ionizes them by a still not understood mechanism. In contrast to ESI, MALDI produces mostly singly charged molecules, and thus simplifies the analysis of complex samples because typically a single spectral band is produced by each species. In addition, since it is not coupled to chromatographic systems, it generally offers higher throughputs. MALDI is being extensively used for the analysis of macromolecules, such as proteins or glycans [33,34].

### 3.3.1 Mass Analyzers

Several designs of mass analyzers exist, each offering different performances in terms of sensitivity, mass accuracy and resolution, mass and concentration dynamic range, or acquisition time. Quadrupoles are the most common kind. In quadrupoles, ion beams transit through an oscillating radiofrequency field that is generated by alternating the voltage on four parallel rods. Depending on the field, only ions with a given mass to charge ( $m/z$ ) ratio cross the quadrupole and reach the detector, whereas all ions that have a different  $m/z$  value are excluded. By acting as a mass filter, quadrupoles exhibit the best characteristics among MS analyzers for the quantitative determination of solute concentrations. The drawback of quadrupoles is that, since only a single  $m/z$  value can be detected at a time point, the duty cycle and hence the sensitivity are inversely proportional to the number of  $m/z$  ratios monitored. A scan over a few hundred amu is typically performed in about 1 second, but more than 99% of the ions are lost during scanning. Detection of a single (or a few)  $m/z$  value by selected ion monitoring (SIM) increases the sensitivity by 2–3 orders of magnitude.

Time-of-flight (TOF) mass analyzers measure the time that ions require to complete a flight path in a vacuum chamber. Because all ions are accelerated in the same electric field, the  $m/z$  ratio can be derived from the elapsed time. In contrast to quadrupoles, TOF detection occurs simultaneously over the entire mass range at faster rates. Notably, TOF MS rapidly detects ions with  $m/z$  up to 100,000–500,000 amu. Mass accuracy and resolution of TOF MS depend on path length and are generally higher than in quadrupoles. Highest performances are obtained in TOF tubes with flight trajectories that are orthogonal to the inlet ion beam, in which a mass accuracy of 10 ppm and resolution of around 10,000 (discriminating 1,000.0 from 1,000.1 amu ions) is routinely achieved.

In addition, since ions are analyzed in packets, TOF MS has a natural compatibility with the pulsed ion release of MALDI interfaces. If there is interest in a given  $m/z$  value, the pulsing frequencies can be adjusted to increase the duty cycle and the sensitivity by about one order of magnitude (lower micromolar range).

A third family of mass analyzers, ion traps, fills a three-dimensional (3D) or linear chamber with ions. Trapped ions are contained by electric fields and, depending on their  $m/z$  ratio, sequentially and selectively excited from the trap by overlying electric fields. At specific voltage settings that correlate with  $m/z$ , resonating ions escape the chamber and are counted by a detector. Linear ion traps can accommodate 40–70 times more ions than 3D ion traps. This correlates with higher duty cycles, sensitivities, and linear ranges. Because of the continuous development of instrumentation, it is rather difficult to compare ion traps and other mass analyzers. Ion traps are generally less rapid, resolving, and accurate than TOF detectors. However, traps are preferred for quantitative purposes, especially for small molecules. An exception is represented by a particular ion trap, the Fourier Transform Ion Cyclotron Resonance (FT-ICR) MS [35], which best supports the identification of unknown compounds owing to its maximum mass resolution (100,000–500,000) and accuracy ( $< 1$  ppm).

### 3.3.2 Applications of MS Detectors

Some MS instruments feature the possibility of selection of an ion, fragmentation by high-energy collisions with gas molecules, and measurement of the mass of the resulting ionic subunits to provide detailed structural information. This so-called MS/MS or tandem MS is obtained either in space or in time. When in space, a (first) quadrupole is operated as a mass filter to let transit only molecules with a given  $m/z$ , which enter a collision chamber (a quadrupole as well) and are then fragmented. The resulting ions are subsequently passed to a second mass analyzer (quadrupole, TOF, or an ion trap) and measured. Tandem MS in time is achieved in ion traps, where the steps of parent ion selection, fragmentation, and fragment detection are performed sequentially within the trap. Using ion traps, this procedure can be repeated several times ( $MS^n$ ) to gain even more detailed insights about the structure of fragments. Three modes are included in MS/MS: product ion, parent ion, and neutral loss scans [36]. The first is used for the structural study of a molecule [37], the resolution of two molecules with identical molecular weight [38], or the identification of proteins and peptides [39]. The latter two modes are employed to detect selectively and quantify, within a complex sample and over a wide  $m/z$  range, only those molecules bearing a given substructure or functional group.

The suitability of MS detectors for quantitative analyses is debated. For example, ESI is a competitive process and, occasionally, matrix (background) material interferes with the ionization of the analyte [40]. These effects of ion suppression are especially aggravated when several species coelute, such as in the case of biological extracts or with direct infusion (without previous separation). Generally, hydrophilic species are more sensitive to ion suppression than hydrophobic ones, which tend to concentrate on the droplet surface during ESI [41]. In MALDI, the sample dispersion is often inhomogeneous or the matrix crystals unevenly distributed on the surface. A truthful representation of the sample composition is obtained exclusively upon thorough laser desorption of the entire spot. In addition, with some instruments the transmission of ions in the mass

analyzers is not ideal and biased in some regions of the mass range because of ion optics when a wide  $m/z$  interval is scanned. After carefully considering these possible error sources, robust concentration estimation with standard deviations of 1–5% are realistic. Internal calibrators at a known concentration can be used to improve the reproducibility of the measurement or to expand the linear dynamic range [42]. Good calibrators are either molecules closely similar to the analyte or, ideally, an analog with stable isotope labels, for example,  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ , or  $^{18}\text{O}$ .

The contribution of MS to identification of compounds and quantification of their concentration is complementary to other detection techniques and, despite being very practical and versatile, it remains fundamentally replaceable. However, knowledge of molecular weight is a prerequisite for techniques that rely on the synergies with stable isotopic tracers. In fact, powerful analytical methods exist to obtain important insights on cell dynamics from the ratiometric measurement of marked and not-marked species (or atoms). We cite, for example, (1) relative abundances of virtually all metabolites or proteins in two separate cultures are quantified based on the isotope dilution theory [43–45]; (2) information on the mechanisms and kinetics of nonlinear chemical processes can be extracted from response tracer experiments [46–47]; and (3) the labeling patterns in metabolic intermediates are used to resolve the relative rate in convergent reactions *in vivo* [48,49].

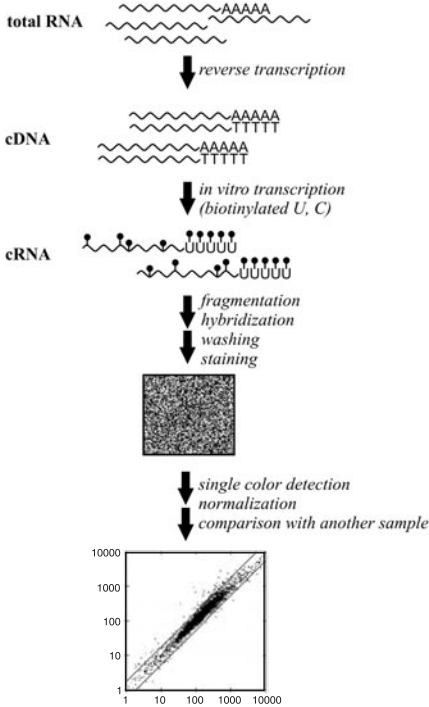
### 3.4 Genome-Wide Analyses of mRNA and Proteins

The central dogma of molecular biology states that DNA-encoded genes define the possible variants of mRNA which, in turn, are meticulously translated into proteins. As these steps progress according to exact rules, it follows that the DNA sequence of a gene can be used to deduce the corresponding protein sequence. However, this conclusion is not completely true. In fact, alternative processing of mRNA occurs in eukaryotic cells, during which fragments of the original sequence are excised (spliced) and a shortened protein is synthesized. Hence, a more correct formulation is that the DNA sequence allows (at least) to predict all possible protein sequences that may be generated from a cell. Despite this enormous number of hypothetical variants, technologies are under development to quantify and characterize the entire mRNA or protein complement of a homogeneous cell population.

#### 3.4.1 DNA Microarrays

Owing to the increasing availability of complete genome sequences, the analysis of transcription by DNA microarrays has managed to achieve an almost genomic scale measurement of mRNA levels. DNA microarrays were initially developed to quantify relative gene expression at the transcriptional level, that is, mRNA. The approach has then been extended to the analysis of single nucleotide polymorphisms (SNP) in DNA [50] or to quantify the abundance of recombinant DNA tags [51]. Compared to alternative technologies, microarrays exhibit the highest throughput with up to  $10^6$  features measured in one single hybridization experiment. A DNA microarray consists of single-stranded oligonucleotides with the sequences of interest that have been imprinted at separated and known locations on a solid surface. Detection of both mRNA and DNA

## (A) High-density arrays



## (B) cDNA microarray

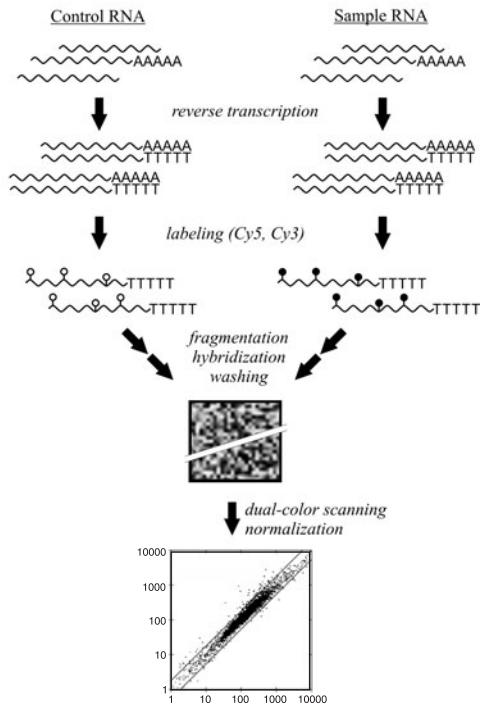


Fig. 3.2 Gene expression analysis with microarrays. cDNA is obtained by reverse transcription of mRNA using oligo(dT) primers. With high-density commercial arrays (A), biotinylated ribonucleotide analogs are incorporated during in vitro transcription of the cDNA. After fragmentation and hybridization, fluorescence detection occurs by streptavidin-mediated binding of fluorophores to biotin. One hybridization is performed per each sample analyzed. Custom-made cDNA arrays (B) enable direct comparison in a single hybridization of two different samples that were previously labeled with distinct fluorophores.

samples relies on the same principle, which is based on the specific binding of strands in a DNA duplex. A labeled DNA complement of mRNA or DNA is synthesized enzymatically in vitro by reverse transcriptase or polymerase chain reaction, respectively. The resulting cDNA is then hybridized on the microarray. Upon wash and detection by confocal fluorescence microscopy, the signal intensity of each spot provides a measure of the corresponding sequence abundance, owing to the highly selective interactions of complementary strands (fig. 3.2).

Essentially two fabrication approaches exist, leading to microarrays with different properties. In one approach, customized cDNA probes are prepared separately and then mechanically spotted on the support, with a density of about  $10,000 \text{ spots/cm}^2$  [52,53]. Because of errors in the handling and printing of probes, this procedure is prone to erroneous arraying [54]. The second approach uses photolithographic techniques to carry out the parallel synthesis of the oligonucleotides directly on the surface [55,56]. The length of the probes is typically for only 14 to 25 bases, much shorter than for

imprinted cDNAs. However, since photolithography enables densities in the range of  $10^6$  spots/cm<sup>2</sup> and higher, multiple different probes are typically spotted for each gene to provide redundancy. Short probes are particularly advantageous in the study of SNPs [50] or differential mRNA splicing [57]. With a customized cDNA array the gene expression in two independent samples can be simultaneously quantified and compared within a single hybridization, when different fluorophores are employed. If only one label is used, as with commercial high-density arrays, intensities are normalized by calculating the total variance in a hybridization of all probes or of a given subset of constitutively expressed genes, that is, those involved in maintenance activities. With fluorescent detection, intensities, and thereby sequences abundance, can be efficiently measured over 2–3 orders of magnitude. Technical and practical details have been described elsewhere [58–60].

Nowadays, microarrays are principally used to reveal differences in gene expression between two (or more) classes of samples and to unravel cellular regulation and adaptation, to investigate gene function, or to identify signature profiles for purposes of classification and diagnostics. In all applications, elaborate bioinformatic tools are required to reduce the complex and large dataset into essential information. Statistical analysis is challenging: beyond the mere analysis of variance within and between duplicates or different samples, noise filtering, and normalization, this task also comprises separating genes causally involved in the primary response from those altered by secondary unspecific mechanisms.

Cross-hybridization hinders precise measurement of mRNAs, especially for rare sequences [61,62]. Also with an optimized probe set that is designed to minimize unspecific hybridization, the typical incubation time of about 12 h dedicated to hybridization does not allow low abundance targets to reach thermodynamic equilibrium, and their concentration is overestimated [61]. It follows that DNA arrays are also not suited for directly measuring expression levels in single cells, perhaps with the exception of enormously sized cells such as oocytes. This intrinsic limitation is obviated by first amplifying the RNA, although some of the ratios between mRNA may change in this additional step [63].

Despite the impressive number of more than 4,000 microarray-based studies published in about a decade, and the proliferation of sophisticated bioinformatic algorithms for the analysis, some criticism is raised by the general reliability and significance of microarray results [64]. Still, the outcomes of such transcript analyses diverge frequently between different RNA isolation procedures, microarray platforms, software, users, and laboratories. The introduction of standard protocols and the automation of sensitive steps contribute to improved overall reproducibility of the entire process. Nevertheless, recent studies emphasize the inherent discrepancy between photolithographic and spotted arrays, between statistical algorithms, and between the set of genes catalogued by the microarrays as altered and that delivered in studies of a different nature [64,65]. At least for quantitative gene transcript analyses, it is advisable to utilize complementary approaches, for example, real-time PCR, to validate the results.

### 3.4.2 Proteome Analyses

Protein function is a direct consequence of its amino acid sequence. Further, protein activity depends on proper folding, post-translational modifications, and interactions

with other species. In analogy to transcript analyses, it is desirable to have technologies that enable the measurement of these characteristics on a genome-wide scale.

To quantify protein levels, 2D polyacrylamide gel electrophoresis is used to separate thousands of proteins from biological samples based on their size and isoelectric point [66]. Abundance is determined by radiometric, colorimetric, or fluorescent detection of the proteins in the gel [67,68]. Commonly used stains are Coomassie Blue, silver stain, and fluorescent dyes. Coomassie Blue stains have limits of detection in the 8–10 ng level and exhibit linear responses over a 10–30-fold range of concentrations. Silver staining is about one order of magnitude more sensitive (0.5 ng) and the linear range spans over a 40–50-fold variation in protein amounts. Best results with silver staining are obtained following aldehyde fixation, but this makes subsequent identification by N-terminal sequencing or MS impossible. Noncovalent fluorescent dyes such as SYPRO Ruby [69] offer the sensitivity of silver stains and linear responses over 2–3 orders of magnitude, and are fully compatible with downstream analyses or other staining procedures [67,70,71]. The stained gels are then digitalized, and the spots identified and quantified by software. With this procedure, gel-to-gel reproducibility in the range of 5–30% can be attained [67,70].

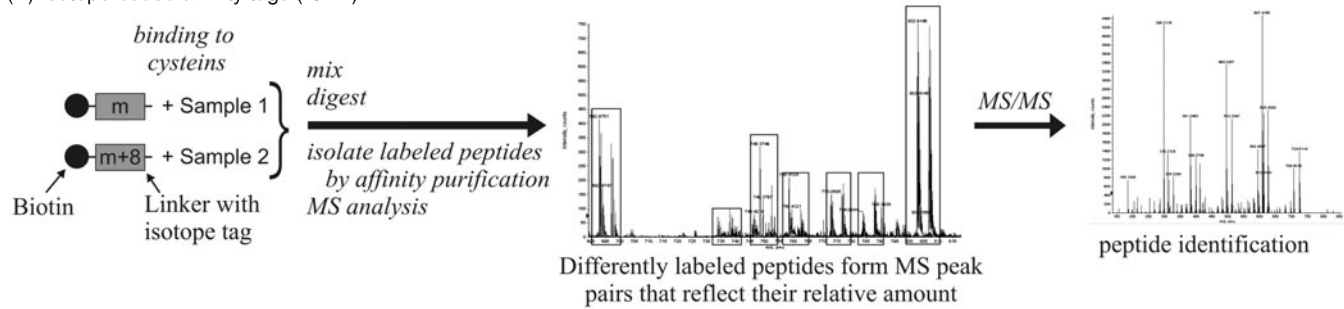
Since binding of dyes depends on the amino acids composition of the protein, secondary modifications, or impurities (e.g., DNA and polysaccharides in the case of silver stain), reliable transformations between spot intensity and protein amount are obtained only after calibrating or carefully accounting for such effects [72]. Beside these general-purpose dyes, special techniques are used to stain only those proteins with selected modifications, for example, phosphoproteins [73], glycoproteins [74], or carbonylated proteins [75]. Identification by MS follows quantification. For this purpose, the protein from a specific spot is extracted from the gel and digested to shorter peptides by a sequence-specific protease such as trypsin. The digest is then analyzed by MS or MS/MS, and the masses of the resulting peptides are searched in a database or, when fragment information is collected in the MS/MS mode, used for *de novo* sequencing [39]. Because of incomplete digestion and unspecific cleavage, the identified peptides usually cover only 10–30% of the entire protein. Notably, MS detection provides access to virtually all post-translational modifications since they lead to modified molecular weights [34,76,77].

Separation and quantification with 2D gels is still a tedious process that requires extensive user intervention to match the spots on different gels. In addition, different protocols have to be employed to solubilize and separate the hydrophobic (membrane) protein fraction [78], and low abundance proteins are detected only when starting with large amounts of material [79]. An alternative to gels is multidimensional HPLC, in which two or more different methods are used in sequence to separate proteins or digested peptides before MS detection [80]. The application of LC-MS systems for quantitative proteomics is still in its early stages and, thus, questionable. To date, accurate LC-MS measurements have been shown only for the determination of protein levels in different samples, as obtained by radiometric MS analyses in combination with isotope tagging techniques [81] (fig. 3.3).

In analogy to DNA arrays, protein microarrays are under development. These devices should enable investigators to quantify expression levels, discover bimolecular interactions, assign functions, and perform structural studies [82,83]. Thus far, only the potential of protein arrays has been validated in pilot studies. Before robust and flexible analytical



(A) Isotope-coded affinity tags (ICAT)



(B) Isobaric tags

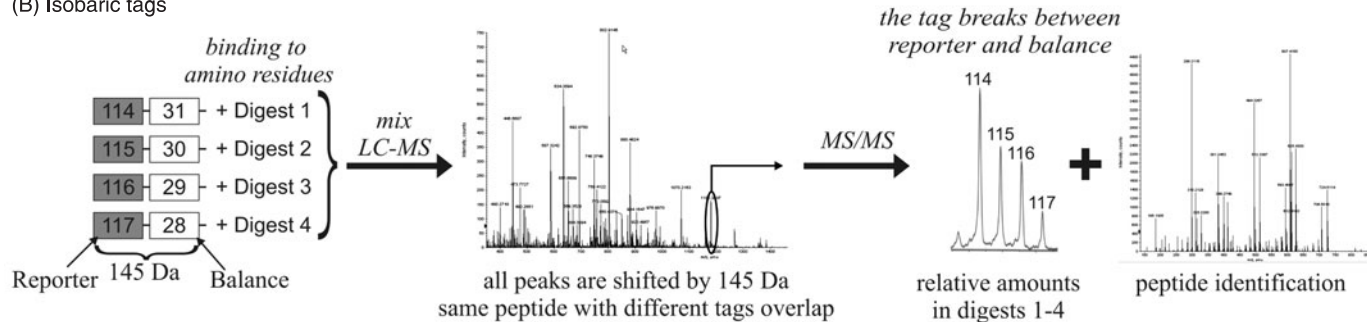


Fig. 3.3 Quantitative proteome analysis by isotopic tags. In (A) isotope-coded affinity tags (ICAT), the cystein residues in protein extracts from two samples are labeled with tags of different molecular weight. Labeled proteins are enriched by affinity purification (between biotin and streptavidin) and quantified with a full MS scan. ICAT is only suited for the analysis of proteins with cystein residues, accounting for about 10% of the total [128]. (B) Isobaric tags are designed to contain one out of four reporters with different molecular weight, but the same total mass due to the balance moiety that compensates for the difference. The tags bind to the amino terminus of peptides. Up to four samples are separately digested and linked to the tags. After mixing, the complexity of the MS spectrum remains unchanged. The relative amounts of a given peptide in the four digests are determined by MS/MS analysis as a result of the relative intensities for the four reporters. Absolute concentrations can be obtained by replacing one of the samples with a standard.

platforms can be established, further technical advances are required to improve general performances and multiplexing capacity.

### 3.5 Fluorescent Imaging

For centuries, light microscopy has constituted the most direct means of monitoring live cell dynamics as they occur. Along with transmitted light observations, fluorescence labeling techniques have generated the greatest advances and successes in the study of the spatial and temporal organization of subcellular structures at a single molecule level. Initially, fluorescent dyes were utilized to localize individual proteins in immobilized and permeabilized dead samples by indirect immunofluorescence, in which an antibody specific for the target protein was used for highly selective recognition. Detection occurs by a second, fluorescently labeled antibody that binds to the constant moiety of the first one [84]. To enable the investigation of structures and dynamics in live cells, a wide range of dyes permeable to the cell membrane have been developed to highlight structural components and, to some extent, cellular functions [69]. However, the permeability of dyes, and thus the general applicability of staining protocols, is affected by the often different composition of the cell membrane and the additional presence of cell walls in nonanimal cells. This problem is partially circumvented by microinjection, at least for those cell classes that can be microinjected and compartments that are accessible. In fact, microinjection is still used to introduce fluorescently labeled oligonucleotides to monitor gene expression in single cells by fluorescence in situ hybridization [85]. In contrast, to track a given protein in living cells with a fluorescent dye, it has to be extracted, purified, tagged with the fluorophore, and re-injected. These procedures are tedious, since proteins with low abundances or those incorporated in complexes are difficult to enrich, and microinjection has to be repeated for each observed cell.

This changed drastically with the cloning of the green fluorescent protein (GFP) from the jellyfish *Aequorea*. The proteic nature of GFP makes it possible to label fluorescently virtually any protein by fusing the GFP-encoding gene to the end of the gene of interest using recombinant DNA techniques. The native gene in the chromosome can be replaced by the tagged form to label automatically and endogenously all the copies of the protein of interest. GFP fusions replace external labeling protocols and, hence, afford the means to study proteins that are difficult to isolate, to apply live imaging in cells that are not suitable for microinjection, or even to track gene expression in whole organisms [86]. Several mutants of GFP have been isolated with different excitation and emission spectra, such as yellow (YFP), cyan (CFP), or blue fluorescence proteins (BFP) [87]. Additionally, a novel red fluorescent protein (RFP) was cloned from *Discosoma* coral [88]. These color variants allow for simultaneous tracking of multiple proteins, quantifying their occurrence in complex structures, or monitoring gene expression levels. Boosted by the unlimited availability of fluorescently tagged proteins and the increasing demand for in vivo analyses, several imaging techniques have been developed. We briefly review their basic principles and applications.

#### 3.5.1 Photobleaching and Photoactivation

With scanning confocal microscopy, a beam of light can be focused on a small area. If such a localized beam is used to excite a fluorophore, this can inactivate it and lead to

an irreversible loss of fluorescence. Upon this so-called photobleaching, fluorescence at the targeted spot can be restored only by replacement of the fluorophore. Therefore, the time for fluorescence recovery after photobleaching (FRAP) is used to infer the turnover of a protein in complexes [89], or estimate diffusion coefficients in cellular compartments [90] and membranes [91]. An extension to FRAP and its related fluorescence loss in photobleaching technique (FLIP) [92] employs two different fluorophores tagged on the same protein. By the selective bleaching of only one fluorophore, this innovation permits the tracking of the localization of bleached proteins over time after the photobleaching event, and to distinguish them from the unbleached fraction by ratiometric measurement.

A proof-of-principle for this method, termed fluorescence localization after photobleaching (FLAP), was provided in experiments with CFP and YFP in the demonstration of the different relocation dynamics between the monomeric and the filamentous forms of actin in living cells [93]. In an alternative to FLAP, similar readouts may be obtained using photoactivation [94], by which the fluorescent emission of the probe is increased upon illumination at a specific wavelength and the activated fluorophores are tracked in time. Photoactivatable GFP variants exhibit a 100-fold increase in fluorescence upon exposure to 400 nm light [87]. Additionally, RFP mutants were isolated with UV-induced green to red fluorescence [95], or a green-induced increase of red fluorescence [96]. Photoactivatable protein can also be obtained using photo-decaging fluorescent tags [97]. Since labeling has to be accomplished externally, this approach is rarely preferred to fluorescent protein fusions. Caged probes have been used, for example, to great advantage in the investigation of phosphorylation-dependent protein–protein interactions [98].

### 3.5.2 Fluorescence Speckle Microscopy (FSM)

FSM is a method for analyzing the assembly, disassembly, and movement of macromolecular structures in vivo (or in vitro) [99,100]. Compared to photobleaching and photoactivation-based approaches, FSM delivers similar information, but on a larger area of the cell. A small fraction of the monomer proteins that make up the assembled polymeric complex is fluorescently labeled and microinjected. Upon incorporation of labeled samples into the polymer, fluorescent speckles become visible by microscopy because of the variable number of fluorophores in each resolution-limited image region. If the relative amounts of labeled and unlabeled monomers are chosen properly depending on the monomer size, fluorescent emission wavelength, and numerical aperture of the microscope, the speckles mark the macromolecules with a discontinuous pattern similar to a barcode. The movement, appearance, and disappearance of speckles, corresponding to movement, polymerization, and depolymerization of the macromolecule, respectively, are tracked over time by wide field or spinning disk confocal fluorescence microscopy to ensure high temporal (5 s) and spatial resolution (0.25  $\mu\text{m}$ ). Image deconvolution and kinetic analysis are then performed by powerful algorithms on the basis of the trajectory from birth to death for each of the many thousands of speckles [101]. FSM has been applied to the study of cytoskeleton dynamics during cell migration, division, and morphogenesis [99,102].

### 3.5.3 Total Internal Reflection Fluorescence Microscopy (TIRFM)

When incident light passes from a medium with a high refractive index to one with a lower index, such as from glass to the aqueous cellular cytosol, total reflection occurs at the interface for incidence angles exceeding a critical value. Despite total reflection, an evanescent field with the same wavelength as the incident light is created in the aqueous phase (the cell) in the proximity of the reflection spot, and decays exponentially in space. In TIRFM, a totally reflected light beam is used to excite selectively those fluorophores that are very close to the coverslip, typically 50–100 nm [103,104]. This technique is optimally suited to explore surface dynamics in live cells at a single molecule resolution. Indeed, it has found important applications in the study of membrane binding, diffusion, and fusion processes [105], endocytosis [106], signaling cascades [107], voltage-gated  $\text{Ca}^{2+}$  channels [108], and viral infection [109].

### 3.5.4 Fluorescence Correlation Spectroscopy (FCS)

A complementary single molecule detection technique is FCS [110,111] (fig. 3.4). Here, confocal microscopy is used to record minute fluorescence fluctuations in a very small

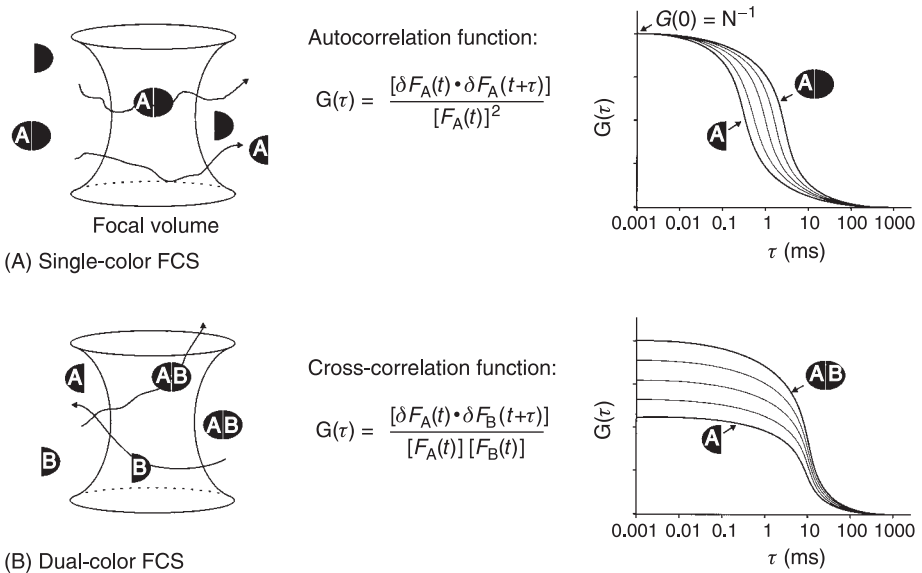


Fig. 3.4 Scheme of fluorescence correlation spectroscopy (FCS). In single (A) and dual (B) color FCS, the fluorescence in the observation volume of one (X) or two (X, Y) distinct classes of fluorophores, respectively, is monitored over time. The correlation functions are then used for the temporal analysis of the signals to determine the characteristic residence times of the labeled probes and their diffusion or binding constants. The correlation plots depict ideal cases.  $G(0)$  is the correlation amplitude, which is inversely proportional to the average number of fluorescent labels simultaneously present in the observation volume ( $N$ ).

(less than 1 femtoliter) and precisely localized observation volume, within compartments or at their boundary on membranes. Intensity fluctuations are caused (1) by changes in the number of fluorophores in the observation volume, induced by diffusive motion, or (2) by variations in the fluorescence emission during the residence of a fluorophore, as a consequence of molecular interactions, or variations in the excitation or in the microenvironment. For reliable analyses, the number of fluorescent molecules in the observation volume has to be kept small (less than 10), which means that fluorophore concentrations have to be in the nanomolar range. A temporal analysis of the fluorescence fluctuations from the average intensity is then performed using an autocorrelation function to estimate the transition time for events affecting the mobility of the labeled molecule such as association and dissociations reactions [112], or to characterize processes affecting the fluorescent emission such as proton transfers [113], conformational changes [114], or oxygen binding [115]. FCS offers subnanomolar sensitivity and dynamic ranges of time that span over 3–4 orders of magnitude [111].

There are, however, limits in the resolution of heterogeneous populations. In the particular case of multiple diffusing fluorescent species, for example in binding studies, the diffusion times must differ by a factor of 1.6–2.0 to be resolved, corresponding to a 4–8-fold increase in molecular weight. This limitation is alleviated in dual-color fluorescence cross-correlation spectroscopy (DC-FCCS) [116], in which the fluorescence fluctuation is simultaneously recorded in the same observation volume for two different species labeled with distinguishable fluorophores. The fraction of double-labeled and hence associated species is derived from the cross-correlation function between the two fluorescences (fig. 3.4B).

### 3.5.5 Fluorescence Resonance Energy Transfer (FRET)

This technique is used to sense the proximity of two fluorescently labeled molecules. The fluorophore couple is chosen to have a donor that in the excited state transfers its energy by dipole–dipole coupling to an acceptor, that is, the second fluorophore, which re-emits the light at a longer wavelength [117]. Among others, FRET pairs with appropriate spectral overlaps are BFP–GFP or CFP–YFP. Because the transfer efficiency correlates with the inverse sixth power of the distance between the fluorophores and depends on their spatial orientation, the experimental detection of FRET is possible only for pairs separated in space by 1–10 nm. If position and orientation of the fluorophore pair are favorable and FRET occurs, the fluorescent emission of the donor is quenched whereas the acceptor begins to fluoresce. The transfer efficiency can be derived by the ratio of the two emissions [118].

Alternative approaches to measure FRET involve photobleaching [119] or the donor's fluorescence lifetime, which is shortened by FRET [120]. Particularly in steady-state FRET experiments, intensity-based FRET techniques are sensitive to cross-talk caused by the partial overlap of the fluorophores' excitation and emission spectra. To obtain exact FRET efficiencies, background fluorescence and FRET-independent variations have to be experimentally quantified for both fluorophores. Two labeling designs are basically available. First, the two fluorophores can be tagged to different molecules to detect their reciprocal binding [121]. Second, both fluorophores are fused to two independent domains of the same molecule, typically a protein, and FRET efficiency depends on conformational changes due to, for example, binding of a ligand. The latter, unimolecular

design has found practical application in the development of FRET-based biosensors for intracellular maltose,  $\text{Ca}^{2+}$ , cAMP, or voltage changes across the membrane [122,123].

### 3.5.6 Future Trends

Fluorescent imaging has already been demonstrated to be a versatile and powerful tool to unravel molecular interactions and dynamics in live cells. The development of more sophisticated techniques is in continuous progress and involves all aspects of fluorescent imaging: fluorophores, instrumentation, and analysis software. The advances are intended to (1) ensure controlled physiological conditions during detection [124]; (2) minimize photodamage of cellular components by decreasing light exposition, scanning time, and unwanted photobleaching, for example by multi-photon excitation [125]; (3) increase spatial resolution beyond diffraction limits [126]; and (4) extend the palette of fluorophores with distinct spectral properties. We expect this ongoing refinement process to further endorse the paramount role of fluorescent imaging in live cell dynamics.

## 3.6 Conclusions

To date, the comprehensive measurement of all species in a single live cell over time remains a vision. Nevertheless, conceptual and technical developments in the last decade have enabled progress in two complementary directions. On the one hand, parallel analyses of thousands of species are possible thanks to mass spectrometry and microarrays. Yet, these technologies have the drawbacks that cells have to be disrupted to extract the analytes, and single-cell analyses are still limited by insufficient sensitivity or unspecific interactions, respectively. On the other hand, real-time imaging techniques deliver precise measurements at a single molecule level over time and with high spatial resolution, but are limited to the observations of only a few species at the time (mostly proteins). In both cases, we expect the bottlenecks to be (partially) relieved in the future.

Overall, it is imperative to ensure that the observed cellular responses are not caused by the measurement itself or by the sampling procedure. The nature, extent, and time frame of such effects must be carefully assessed to avoid misleading artifacts.

## References

- [1] Sawada, S.; Ono, R.; Sato, T.; Suzuki, S.; Arakawa, O.; Kasai, M. Determination of sugar phosphates and nucleotides related to photosynthetic metabolism by high-performance anion-exchange liquid chromatography with fluorometric and ultraviolet detection. *Anal. Biochem.* **2003**, *314*, 63–69.
- [2] Vogt, A. M.; Ackermann, C.; Noe, T.; Jensen, D.; Kübler, W. Simultaneous detection of high energy phosphates and metabolites of glycolysis and the Krebs cycle by HPLC. *Biochem. Biophys. Res. Commun.* **1998**, *248*, 527–532.
- [3] Bhattacharya, M.; Fuhrman, L.; Ingram, A.; Nickerson, K. W.; Conway, T. Single-run separation and detection of multiple metabolic intermediates by anion-exchange high-performance liquid chromatography and application to cell pool extracts prepared from *Escherichia coli*. *Anal. Biochem.* **1995**, *232*, 98–106.

- [4] Taha, T. S.; Deits, T. L. Detection of metabolites of the Entner-Doudoroff pathway by HPLC with pulsed amperometry: application to assays for pathway enzymes. *Anal. Biochem.* **1994**, *219*, 115–120.
- [5] Huck, J. H.; Struys, E. A.; Verhoeven, N. M.; Jakobs, C.; van der Knaap, M. S. Profiling of pentose phosphate pathway intermediates in blood spots by tandem mass spectrometry: application to transaldolase deficiency. *Clin. Chem.* **2003**, *49*, 1375–1380.
- [6] van Dam, J. C.; Eman, M. R.; Frank, J.; Lange, H. C.; van Dedem, G. W.; Heijnen, J. J. Analysis of glycolytic intermediates in *Saccharomyces cerevisiae* using anion exchange chromatography and electrospray ionization with tandem mass spectrometric detection. *Anal. Chim. Acta* **2002**, *460*, 209–218.
- [7] Ohie, T.; Fu, X.; Iga, M.; Kimura, M.; Yamaguchi, S. Gas chromatography-mass spectrometry with *tert*-butyldimethylsilyl derivation: use of the simplified sample preparations and the automated data system to screen for organic acidemias. *Chromatogr. B* **2000**, *746*, 63–73.
- [8] Kim, K. R.; Hahn, M. K.; Zlatkis, A.; Horning, E. C.; Middleditch, B. S. Simultaneous gas chromatography of volatile and non-volatile carboxylic acids as *tert*-butyldimethylsilyl derivatives. *J. Chromatogr.* **1989**, *468*, 289–301.
- [9] Roessner, U.; Wagner, C.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. Technical advance: simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J.* **2000**, *23*, 131–142.
- [10] Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **2000**, *18*, 1157–1161.
- [11] von Roepenack-Lahaye, E.; Degenkolb, T.; Zerjeski, M.; Franz, M.; Roth, U.; Wessjohann, L.; Schmidt, J.; Scheel, D.; Clemens, S. Profiling of *Arabidopsis* secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. *Plant. Physiol.* **2004**, *134*, 548–559.
- [12] Tolstikov, V. V.; Lommen, A.; Nakanishi, K.; Tanaka, N.; Fiehn, O. Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics. *Anal. Chem.* **2003**, *75*, 6737–6740.
- [13] Tolstikov, V. V.; Fiehn, O. Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal. Biochem.* **2002**, *301*, 298–307.
- [14] Schwartz, H. E.; Ulfelder, K. J.; Chen, F. T. A.; Pentoney, S. L. The utility of laser-induced fluorescence detection in applications of capillary electrophoresis. *J. Cap. Elec.* **1994**, *2*, 36–54.
- [15] Moini, M. Capillary electrophoresis mass spectrometry and its application to the analysis of biological mixtures. *Anal. Bioanal. Chem.* **2002**, *373*, 466–480.
- [16] Perrett, D. Capillary electrophoresis in clinical chemistry. *Ann. Clin. Biochem.* **1999**, *36*, 133–150.
- [17] Woods, L. A.; Ewing, A. G. Analysis of single mammalian cells with capillary electrophoresis. *Anal. Bioanal. Chem.* **2003**, *376*, 281–283.
- [18] Chen, S.; Lillard, S. J. Continuous cell introduction for the analysis of individual cells by capillary electrophoresis. *Anal. Chem.* **2001**, *73*, 111–118.
- [19] Zhang, H.; Jin, W. Determination of different forms of human interferon-gamma in single natural killer cells by capillary electrophoresis with on-capillary immunoreaction and laser-induced fluorescence detection. *Electrophoresis* **2004**, *25*, 1090–1095.
- [20] Lillard, S. J.; Yeung, E. S.; Lautamo, R. M.; Mao, D. T. Separation of hemoglobin variants in single human erythrocytes by capillary electrophoresis with laser-induced native fluorescence detection. *J. Chromatogr. A* **1995**, *718*, 397–404.
- [21] Zhang, H.; Jin, W. Analysis of amino acids in individual human erythrocytes by capillary electrophoresis with electroporation for intracellular derivatization and laser-induced fluorescence detection. *Electrophoresis* **2004**, *25*, 480–486.
- [22] Xue, Q.; Yeung, E. S. Indirect fluorescence determination of lactate and pyruvate in single erythrocytes by capillary electrophoresis. *J. Chromatogr. A* **1994**, *661*, 287–295.
- [23] Bossuyt, X. Separation of serum proteins by automated capillary zone electrophoresis. *Clin. Chem. Lab. Med.* **2003**, *41*, 762–772.

- [24] Righetti, P. G. Capillary electrophoretic analysis of proteins and peptides of biomedical and pharmacological interest. *Biopharm. Drug Dispos.* **2001**, 22, 337–351.
- [25] Soga, T.; Ohashi, Y.; Ueno, Y.; Naraoka, H.; Tomita, M.; Nishioka, T. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J. Proteome Res.* **2003**, 2, 488–494.
- [26] Nishi, H.; Fukuyama, T.; Terabe, T. Chiral separation by cyclodextrin-modified micellar electrokinetic chromatography. *J. Chromatogr.* **1991**, 553, 503–516.
- [27] Heegaard, N. H. Applications of affinity interactions in capillary electrophoresis. *Electrophoresis* **2003**, 24, 3879–3891.
- [28] Gross, J. H. *Mass Spectrometry: A Textbook*; Springer: New York, 2004.
- [29] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, 246, 64–71.
- [30] Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y. Protein and polymer analyses up to  $m/z$  100,000 by laser ionization time-of-flight mass spectrometry. *Rapid. Commun. Mass Spectrom.* **1988**, 2, 151–153.
- [31] Karas, M.; Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **1988**, 60, 2299–2301.
- [32] Cole, R. B., Ed. *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation, and Applications*; Wiley: New York, 1997.
- [33] Marvin, L. F.; Roberts, M. A.; Fay, L. B. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry. *Clin. Chim. Acta* **2003**, 337, 11–21.
- [34] Areces, L. B.; Matafora, V.; Bachi, A. Analysis of protein phosphorylation by mass spectrometry. *Eur. J. Mass Spectrom.* **2004**, 10, 383–392.
- [35] Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **1998**, 17, 1–35.
- [36] Schwartz, J. C.; Wade, A. P.; Enke, C. G.; Cooks, R. G. Systematic delineation of scan modes in multidimensional mass spectrometry. *Anal. Chem.* **1990**, 62, 1809–1818.
- [37] Griffiths, W. J. Tandem mass spectrometry in the study of fatty acids, bile acids, and steroids. *Mass Spectrom. Rev.* **2003**, 22, 81–152.
- [38] Leavell, M. D.; Leary, J. A. Probing isomeric differences of phosphorylated carbohydrates through the use of ion/molecule reactions and FT-ICR MS. *J. Am. Soc. Mass Spectrom.* **2003**, 14, 323–331.
- [39] Chapman, J. R., Ed. *Mass Spectrometry of Proteins and Peptides*; Humana Press: Totowa, NJ, 2000.
- [40] Choi, B. K.; Hercules, D. M.; Gusev, A. I. Effect of liquid chromatography separation of complex matrices on liquid chromatography-tandem mass spectrometry signal suppression. *J. Chromatogr. A* **2001**, 907, 337–342.
- [41] Bonfiglio, R.; King, R. C.; Olah, T. V.; Merkle, K. The effects of sample preparation methods on the variability of the electrospray ionization response for model drug compounds. *Rapid Commun. Mass Spectrom.* **1999**, 13, 1175–1185.
- [42] Shi, G. Application of co-eluting structural analog internal standards for expanded linear dynamic range in liquid chromatography/electrospray mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, 17, 202–206.
- [43] Zhou, H.; Ranish, J. A.; Watts, J. D.; Aebersold, R. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nat. Biotechnol.* **2002**, 20, 512–515.
- [44] de Leenheer, A. P.; Thienpont, L. M. Applications of isotope dilution-mass spectrometry in clinical chemistry, pharmacokinetics and toxicology. *Mass Spectrom. Rev.* **1992**, 11, 249–307.
- [45] Mashego, M. R.; Wu, L.; Van Dam, J. C.; Ras, C.; Vinke, J. L.; Van Winden, W. A.; Van Gulik, W. M.; Heijnen, J. J. MIRACLE: mass isotopomer ratio analysis of U- $^{13}\text{C}$ -labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites. *Biotechnol. Bioeng.* **2004**, 85, 620–628.
- [46] Vlad, M. O.; Arkin, A.; Ross, J. Response experiments for nonlinear systems with application to reaction kinetics and genetics. *Proc. Natl. Acad. Sci. USA* **2004**, 101, 7223–7228.



- [47] Vlad, M. O.; Ross, J.; Moran, F.; Joel, Y. Delayed response in tracer experiments and fragment-carrier approach to transit time distribution in nonlinear chemical kinetics. *Int. J. Bif. Chaos* **2002**, *12*, 2599–2618.
- [48] Szyperski, T. Biosynthetically directed fractional  $^{13}\text{C}$ -labeling of proteinogenic amino acids: an efficient analytical tool to investigate intermediary metabolism. *Eur. J. Biochem.* **1995**, *232*, 433–448.
- [49] Fischer, E.; Sauer, U. Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. *Eur. J. Biochem.* **2003**, *270*, 880–891.
- [50] Oefner, P. J. Sequence variation and the biological function of genes: methodological and biological considerations. *J. Chromatogr. B* **2002**, *782*, 3–25.
- [51] Giaever, G.; Flaherty, P.; Kumm, J.; Proctor, M.; Nislow, C.; Jaramillo, D. F.; Chu, A. M.; Jordan, M. I.; Arkin, A. P.; Davis, R. W. Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 793–798.
- [52] Schena, M.; Shalon, D.; Davis, R. W.; Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **1995**, *270*, 467–470.
- [53] Ferguson, J. A.; Boles, T. C.; Adams, C. P.; Walt, D. R. A fiber-optic DNA biosensor microarray for the analysis of gene expression. *Nat. Biotechnol.* **1996**, *14*, 1681–1684.
- [54] Halgren, R. G.; Fielden, M. R.; Fong, C. J.; Zacharewski, T. R. Assessment of clone identity and sequence fidelity for 1189 IMAGE cDNA clones. *Nucleic Acids Res.* **2001**, *29*, 582–588.
- [55] Fodor, S. P.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Lu, A. T.; Solas, D. Light-directed, spatially addressable parallel chemical synthesis. *Science* **1991**, *251*, 767–773.
- [56] McGall, G. H.; Fidanza, F. A. Photolithographic synthesis of arrays. In *Methods in Molecular Biology: DNA Arrays, Methods and Protocols*; Rampal, J. B., Ed.; Humana Press: Totowa, NJ, 2001; Vol. 170, pp 71–101.
- [57] Johnson, J. M.; Castle, J.; Garrett-Engle, P.; Kan, Z.; Loerch, P. M.; Armour, C. D.; Santos, R.; Schadt, E. E.; Stoughton, R.; Shoemaker, D. D. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **2003**, *302*, 2141–2144.
- [58] Rampal, J. B., Ed. *Methods in Molecular Biology: DNA Arrays, Methods and Protocols*; Humana Press: Totowa, NJ, 2001; Vol. 170, pp 264.
- [59] Schena, M., Ed. *DNA Microarrays: A Practical Approach*; Oxford University Press: Oxford, 1999; pp 210.
- [60] Heller, M. J. DNA microarray technology: devices, systems, and applications. *Annu. Rev. Biomed. Eng.* **2002**, *4*, 129–153.
- [61] Bhanot, G.; Louzoun, Y.; Zhu, J.; DeLisi, C. The importance of thermodynamic equilibrium for high throughput gene expression arrays. *Biophys. J.* **2003**, *84*, 124–135.
- [62] Tu, Y.; Stolovitzky, G.; Klein, U. Quantitative noise analysis for gene expression microarray experiments. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14031–14036.
- [63] Kawasaki, E. S. Microarrays and the gene expression profile of a single cell. *Ann. N.Y. Acad. Sci.* **2004**, *1020*, 92–100.
- [64] Miklos, G. L.; Maleszka, R. Microarray reality checks in the context of a complex disease. *Nat. Biotechnol.* **2004**, *22*, 615–621.
- [65] Etienne, W.; Meyer, M. H.; Peppers, J.; Meyer, R. A., Jr. Comparison of mRNA gene expression by RT-PCR and DNA microarray. *Biotechniques* **2004**, *36*, 618–626.
- [66] Wildgruber, R.; Harder, A.; Obermaier, C.; Boguth, G.; Weiss, W.; Fey, S. J.; Larsen, P. M.; Görg, A. Towards higher resolution: two-dimensional electrophoresis of *Saccharomyces cerevisiae* proteins using overlapping narrow immobilized pH gradients. *Electrophoresis* **2000**, *21*, 2610–2616.
- [67] Patton, W. F. Detection technologies in proteome analysis. *J. Chromatogr. B* **2002**, *771*, 3–31.
- [68] Roegerer, J.; Lutter, P.; Reinhardt, R.; Blüggel, M.; Meyer, H. E.; Anselmetti, D. Ultra-sensitive detection of unstained proteins in acrylamide gels by native UV fluorescence. *Anal. Chem.* **2003**, *75*, 157–159.
- [69] Haugland, R. P. *Handbook of Fluorescent Probes and Research Products*, 9th ed.; Molecular Probes: Eugene, OR, 2002.

- [70] Nishihara, J. C.; Champion, K. M. Quantitative evaluation of proteins in one- and two-dimensional polyacrylamide gels using a fluorescent stain. *Electrophoresis* **2002**, *23*, 2203–2215.
- [71] Lauber, W. M.; Carroll, J. A.; Dufield, D. R.; Kiesel, J. R.; Radabaugh, M. R.; Malone, J. P. Mass spectrometry compatibility of two-dimensional gel protein stains. *Electrophoresis* **2001**, *22*, 906–918.
- [72] Fiévet, J.; Dillmann, C.; Lagniel, G.; Davanture, M.; Negroni, L.; Labarre, J.; De Vienne, D. Assessing factors for reliable quantitative proteomics based on two-dimensional gel electrophoresis. *Proteomics* **2004**, *4*, 1939–1949.
- [73] Steinberg, T. H.; Agnew, B. J.; Gee, K. R.; Leung, W. Y.; Goodman, T.; Schulenberg, B.; Hendrickson, J.; Beechem, J. M.; Haugland, R. P.; Patton, W. F. Global quantitative phosphoprotein analysis using Multiplexed Proteomics technology. *Proteomics* **2003**, *3*, 1128–1144.
- [74] Hart, C.; Schulenberg, B.; Steinberg, T. H.; Leung, W. Y.; Patton, W. F. Detection of glycoproteins in polyacrylamide gels and on electroblots using Pro-Q Emerald 488 dye, a fluorescent periodate Schiff-base stain. *Electrophoresis* **2003**, *24*, 588–598.
- [75] Yoo, B. S.; Regnier, F. E. Proteomic analysis of carbonylated proteins in two-dimensional gel electrophoresis using avidin-fluorescein affinity staining. *Electrophoresis* **2004**, *25*, 1334–1341.
- [76] Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **2003**, *21*, 660–666.
- [77] Person, M. D.; Monks, T. J.; Lau, S. S. An integrated approach to identifying chemically induced posttranslational modifications using comparative MALDI-MS and targeted HPLC-ESI-MS/MS. *Chem. Res. Toxicol.* **2003**, *16*, 598–608.
- [78] Molloy, M. P.; Phadke, N. D.; Maddock, J. R.; Andrews, P. C. Two-dimensional electrophoresis and peptide mass fingerprinting of bacterial outer membrane proteins. *Electrophoresis* **2001**, *22*, 1686–1696.
- [79] Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 9390–9395.
- [80] Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., III. Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999**, *17*, 676–682.
- [81] Lee, H.; Yi, E. C.; Wen, B.; Reilly, T. P.; Pohl, L.; Nelson, S.; Aebersold, R.; Goodlett, D. R. Optimization of reversed-phase microcapillary liquid chromatography for quantitative proteomics. *J. Chromatogr. B* **2004**, *803*, 101–110.
- [82] Cutler, P. Protein arrays: the current state-of-the-art. *Proteomics* **2003**, *3*, 3–18.
- [83] Melton, L. Protein arrays: proteomics in multiplex. *Nature* **2004**, *429*, 101–107.
- [84] Hinkle, P. M.; Puskas, J. A. Detection of G protein-coupled receptors by immunofluorescence microscopy. *Methods Mol. Biol.* **2004**, *237*, 127–134.
- [85] Levsky, J. M.; Singer, R. H. Fluorescence in situ hybridization: past, present and future. *J. Cell Sci.* **2003**, *116*, 2833–2838.
- [86] Hadjantonakis, A. K.; Nagy, A. The color of mice: in the light of GFP-variant reporters. *Histochem. Cell Biol.* **2001**, *115*, 49–58.
- [87] Lippincott-Schwartz, J.; Patterson, G. H. Development and use of fluorescent protein markers in living cells. *Science* **2003**, *300*, 87–91.
- [88] Fradkov, A. F.; Chen, Y.; Ding, L.; Barsova, E. V.; Matz, M. V.; Lukyanov, S. A. Novel fluorescent protein from *Discosoma* coral and its mutants possesses a unique far-red fluorescence. *FEBS Lett.* **2000**, *479*, 127–130.
- [89] Stenoien, D. L.; Mielke, M.; Mancini, M. A. Intracellular ataxin1 inclusions contain both fast- and slow-exchanging components. *Nat. Cell Biol.* **2002**, *4*, 806–810.
- [90] Papadopoulos, S.; Endeward, V.; Revesz-Walker, B.; Jurgens, K. D.; Gros, G. Radial and longitudinal diffusion of myoglobin in single living heart and skeletal muscle cells. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5904–5909.
- [91] Mullineaux, C. W. FRAP analysis of photosynthetic membranes. *J. Exp. Bot.* **2004**, *55*, 1207–1211.

- [92] Lippincott-Schwartz, J.; Altan-Bonnet, N.; Patterson, G. H. Photobleaching and photo-activation: following protein dynamics in living cells. *Nat. Cell. Biol.* **2003**, Suppl., S7–S14.
- [93] Dunn, G. A.; Dobbie, I. M.; Monypenny, J.; Holt, M. R.; Zicha, D. Fluorescence localization after photobleaching (FLAP): a new method for studying protein dynamics in living cells. *J. Microsc.* **2002**, 205, 109–112.
- [94] Patterson, G. H.; Lippincott-Schwartz, J. Selective photolabeling of proteins using photoactivatable GFP. *Methods* **2004**, 32, 445–450.
- [95] Ando, R.; Hama, H.; Yamamoto-Hino, M.; Mizuno, H.; Miyawaki, A. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc. Natl. Acad. Sci. USA* **2002**, 99, 12651–12656.
- [96] Chudakov, D. M.; Belousov, V. V.; Zaraisky, A. G.; Novoselov, V. V.; Staroverov, D. B.; Zorov, D. B.; Lukyanov, S.; Lukyanov, K. A. Kindling fluorescent proteins for precise in vivo photolabeling. *Nat. Biotechnol.* **2003**, 21, 191–194.
- [97] Mitchison, T. J.; Sawin, K. E.; Theriot, J. A.; Gee, K.; Mallavarapu, A. Caged fluorescent probes. *Methods Enzymol.* **1998**, 291, 63–78.
- [98] Vazquez, M. E.; Nitz, M.; Stehn, J.; Yaffe, M. B.; Imperiali, B. Fluorescent caged phosphoserine peptides as probes to investigate phosphorylation-dependent protein associations. *J. Am. Chem. Soc.* **2003**, 125, 10150–10151.
- [99] Waterman-Storer, C. M.; Danuser, G. New directions for fluorescent speckle microscopy. *Curr. Biol.* **2002**, 12, R633–R640.
- [100] Danuser, G.; Waterman-Storer, C. M. Quantitative fluorescent speckle microscopy: where it came from and where it is going. *J. Microsc.* **2003**, 211, 191–207.
- [101] Ponti, A.; Vallotton, P.; Salmon, W. C.; Waterman-Storer, C. M.; Danuser, G. Computational analysis of F-actin turnover in cortical actin meshworks using fluorescent speckle microscopy. *Biophys. J.* **2003**, 84, 3336–3352.
- [102] Vallotton, P.; Ponti, A.; Waterman-Storer, C. M.; Salmon, E. D.; Danuser, G. Recovery, visualization, and analysis of actin and tubulin polymer flow in live cells: a fluorescent speckle microscopy study. *Biophys. J.* **2003**, 85, 1289–1306.
- [103] Axelrod, D. Total internal reflection fluorescence microscopy in cell biology. *Methods Enzymol.* **2003**, 361, 1–33.
- [104] Sako, Y.; Uyemura, T. Total internal reflection fluorescence microscopy for single-molecule imaging in living cells. *Cell. Struct. Funct.* **2002**, 27, 357–365.
- [105] Fix, M.; Melia, T. J.; Jaiswal, J. K.; Rappoport, J. Z.; You, D.; Sollner, T. H.; Rothman, J. E.; Simon, S. M. Imaging single membrane fusion events mediated by SNARE proteins. *Proc. Natl. Acad. Sci. USA* **2004**, 101, 7311–7316.
- [106] Rappoport, J.; Simon, S.; Benmerah, A. Understanding living clathrin-coated pits. *Traffic* **2004**, 5, 327–337.
- [107] Sako, Y.; Ichinose, J.; Morimatsu, M.; Ohta, K.; Uyemura, T. Optical bioimaging: from living tissue to a single molecule. Single-molecule visualization of cell signaling processes of epidermal growth factor receptor. *J. Pharmacol. Sci.* **2003**, 93, 253–258.
- [108] Demuro, A.; Parker, I. Imaging the activity and localization of single voltage-gated  $\text{Ca}^{2+}$  channels by total internal reflection fluorescence microscopy. *Biophys. J.* **2004**, 86, 3250–3259.
- [109] Seisenberger, G.; Ried, M. U.; Endress, T.; Buning, H.; Hallek, M.; Brauchle, C. Real-time single-molecule imaging of the infection pathway of an adeno-associated virus. *Science* **2001**, 294, 1929–1932.
- [110] Hess, S. T.; Huang, S.; Heikal, A. A.; Webb, W. W. Biological and chemical applications of fluorescence correlation spectroscopy: a review. *Biochemistry* **2002**, 41, 697–705.
- [111] Kim, S. A.; Schwille, P. Intracellular applications of fluorescence correlation spectroscopy: prospects for neuroscience. *Curr. Opin. Neurobiol.* **2003**, 13, 583–590.
- [112] Briddon, S. J.; Middleton, R. J.; Cordeaux, Y.; Flavin, F. M.; Weinstein, J. A.; George, M. W.; Kellam, B.; Hill, S. J. Quantitative analysis of the formation and diffusion of A1-adenosine receptor-antagonist complexes in single living cells. *Proc Natl. Acad. Sci. USA* **2004**, 101, 4673–4678.

- [113] Metzger, F.; Repunte-Canonigo, V.; Matsushita, S.; Akemann, W.; Diez-Garcia, J.; Ho, C. S.; Iwasato, T.; Grandes, P.; Itohara, S.; Joho, R. H.; Knopfel, T. Transgenic mice expressing a pH and Cl<sup>−</sup> sensing yellow-fluorescent protein under the control of a potassium channel promoter. *Eur. J. Neurosci.* **2002**, *15*, 40–50.
- [114] Schwille, P.; Kummer, S.; Heikal, A. A.; Moerner, W. E.; Webb, W. W. Fluorescence correlation spectroscopy reveals fast optical excitation-driven intramolecular dynamics of yellow fluorescent proteins. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 151–156.
- [115] Lippitz, M.; Erker, W.; Decker, H.; van Holde, K. E.; Basche, T. Two-photon excitation microscopy of tryptophan-containing proteins. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2772–2777.
- [116] Schwille, P.; Meyer-Almes, F. J.; Rigler, R. Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution. *Biophys. J.* **1997**, *72*, 1878–1886.
- [117] Lakowicz, J. R. Energy transfer. In *Principles of Fluorescence Spectroscopy*; Plenum: New York, 1999; pp 367–394.
- [118] Berney, C.; Danuser, G. FRET or no FRET: a quantitative comparison. *Biophys. J.* **2003**, *84*, 3992–4010.
- [119] Karpova, T. S.; Baumann, C. T.; He, L.; Wu, X.; Grammer, A.; Lipsky, P.; Hager, G. L.; McNally, J. G. Fluorescence resonance energy transfer from cyan to yellow fluorescent protein detected by acceptor photobleaching using confocal microscopy and a single laser. *J. Microsc.* **2003**, *209*, 56–70.
- [120] Elangovan, M.; Day, R. N.; Periasamy, A. Nanosecond fluorescence resonance energy transfer-fluorescence lifetime imaging microscopy to localize the protein interactions in a single living cell. *J. Microsc.* **2002**, *205*, 3–14.
- [121] Vilardaga, J. P.; Bunemann, M.; Krasel, C.; Castro, M.; Lohse, M. J. Measurement of the millisecond activation switch of G protein-coupled receptors in living cells. *Nat. Biotechnol.* **2003**, *21*, 807–812.
- [122] Fehr, M.; Frommer, W. B.; Lalonde, S. Visualization of maltose uptake in living yeast cells by fluorescent nanosensors. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9846–9851.
- [123] Miyawaki, A. Fluorescence imaging of physiological activity in complex systems using GFP-based probes. *Curr. Opin. Neurobiol.* **2003**, *13*, 591–596.
- [124] Stephens, D. J.; Allan, V. J. Light microscopy techniques for live cell imaging. *Science* **2003**, *300*, 82–86.
- [125] Xu, C.; Zipfel, W.; Shear, J. B.; Williams, R. M.; Webb, W. W. Multiphoton fluorescence excitation: new spectral windows for biological nonlinear microscopy. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 10763–10768.
- [126] Hell, S. W. Toward fluorescence nanoscopy. *Nat. Biotechnol.* **2003**, *21*, 1347–1355.
- [127] Soga, T.; Ueno, Y.; Naraoka, H.; Ohashi, Y.; Tomita, M.; Nishioka, T. Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. *Anal. Chem.* **2002**, *74*, 2233–2239.
- [128] Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, *17*, 994–999.

## 4

# Computations by Means of Macroscopic Chemical Kinetics

---

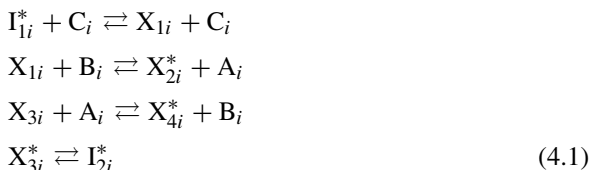
The topic of this chapter may seem like a digression from methods and approaches to reaction mechanisms, but it is not; it is an introduction to it. We worked on both topics for some time and there is a basic connection. Think of an electronic device and ask: how are the logic functions of this device determined? Electronic inputs (voltages and currents) are applied and outputs are measured. A truth table is constructed and from this table the logic functions of the device, and at times some of its components, may be inferred. The device is not subjected to the approach toward a chemical mechanism described in the previous chapter, of taking the device apart and testing its simplest components. (That may have to be done sometimes but is to be avoided if possible.)

Can such an approach be applicable to chemical systems? We show this to be the case by discussing the implementation of logic and computational devices, both sequential machines such as a universal Turing machine (hand computers, laptops) and parallel machines, by means of macroscopic kinetics [1–5]; by giving a brief comparison with neural networks [5,6]; by showing the presence of such devices in chemical and biochemical reaction systems [7]; and by presenting some confirming experiments [8,9]. The next step is clear: if macroscopic chemical kinetics can carry out these electronic functions, then there are likely to be new approaches possible for the determination of complex reaction mechanisms, analogs of such determinations for electronic components.

The discussion in the remainder of this chapter is devoted to illustrations of these topics; it can be skipped, except the last paragraph, without loss of continuity with chapter 5 and beyond.

## 4.1 Chemical Neurons and Logic Gates

A neuron is either on or off depending on the signals it has received. A chemical neuron [10–13] is a similar device. Consider the hypothetical reaction mechanism (fig. 4.1) with the following elementary reaction steps [1]:



Starred species are held constant by buffering, or reservoirs, or flows. A biological example will be given shortly. The macroscopic rate equations are given by

$$\begin{aligned}
 J_{1i} &= k_1 C_i - k_{-1} C_i X_{1i} \\
 J_{2i} &= k_2 X_{1i} B_i - k_{-2} A_i \\
 J_{3i} &= k_3 X_{3i} A_i - k_{-3} B_i \\
 J_{4i} &= k_4 X_{3i} - k_{-4} \\
 dX_{1i}/dt &= J_{1i} - J_{2i} \\
 dX_{3i}/dt &= -J_{3i} - J_{4i} \\
 dA_i/dt &= -d B_i/dt = J_{2i} - J_{3i}
 \end{aligned} \tag{4.2}$$

There is constant inflow of  $I_1$  and a constant outflow of  $X_{2i}$ . The stationary states of this reaction system, obtained by setting the time derivatives of the intermediates equal to zero, are plotted in fig. 4.2. For small values of the catalyst concentration  $C_i$  the concentration of  $A_i$  is small and at  $C_i = 1$  the value of  $A_i$  increases abruptly and remains large thereafter. (The numbers are arbitrary. We use roman letters for species and italic letters for concentrations.) The reaction mechanism represents a chemical neuron, either on or off, as some parameter changes.

Next we consider the construction of logic gates by coupling of chemical neurons [2]. For each neuron in the network there is one copy of the reaction mechanism (fig. 4.1); the neurons are mechanistically similar but chemically distinct. In fig. 4.3 we show a possible coupling of neuron  $j$  to neuron  $i$ . We choose an enzyme mechanism in which the concentration of either  $A_j$  or  $B_j$  becomes an activator or an inhibitor of an enzyme

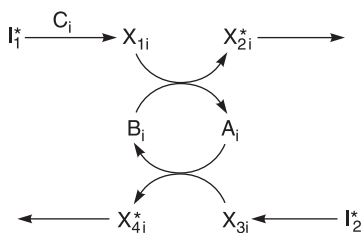


Fig. 4.1 Schematic of reaction mechanism with elementary steps given in eq. (4.1) and mass action rate expressions in eq. (4.2).

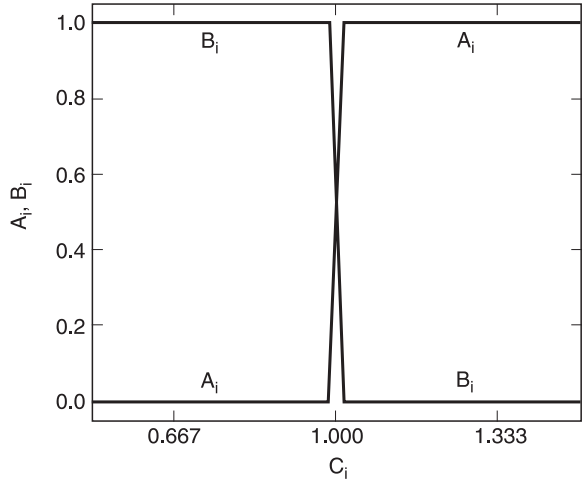


Fig. 4.2 Stationary states of reactants  $A_i$  and  $B_i$  as a function of the catalyst concentration  $C_i$  for the reaction mechanism in fig. 4.1. (From [1].)

$E_{ij}$  for  $C_{ij}$ . Suppose that the state of the neuron  $j$  is to excite neuron  $i$ . We have

$$E_{ij} + A_j = C_{ij} \tag{4.3}$$

The total enzyme concentration  $E_{ij}^0$  is

$$E_{ij}^0 = E_{ij} + C_{ij} \tag{4.4}$$

With the equilibrium constant of the rapid binding of the enzyme and  $A_j$ ,

$$K = \frac{C_{ij}}{E_{ij}A_j} \tag{4.5}$$

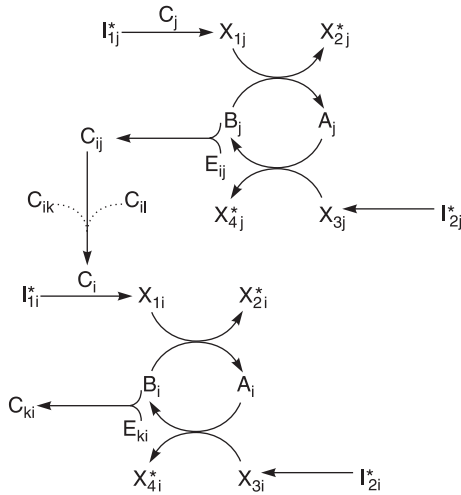


Fig. 4.3 Schematic of two reaction mechanisms constituting neurons  $i$  and  $j$ , and the influence of neurons  $j$ ,  $k$ , and  $l$  on  $i$ . All reactions are reversible, and the concentration of the species marked by  $*$  is held constant. The firing of neuron  $j$  inhibits the firing of neuron  $l$ , and neurons  $k$  and  $l$  (data not shown) also influence the state of neuron  $i$ . The firing of neuron  $i$  inhibits the firing on neuron  $k$ . (From [1].)

we can solve for  $C_{ij}$ :

$$C_{ij} = \frac{E_{ij}^0}{1 + \frac{1}{KA_j}} \quad (4.6)$$

The contributions of all neurons acting on neuron  $i$  are additive and hence the total concentration of the enzyme  $C_i$  is

$$C_i = \sum_j C_{ij} \quad (4.7)$$

For neurons  $j$  and  $k$  to perform an AND operation on neuron  $i$  we need an activation reaction, eq. (4.3), for each of the species  $A_j$  and  $A_k$ . The concentration of the catalyst  $C_i$  is then the sum of  $C_{ij}$  and  $C_{ik}$  with the result, for  $K = E_{ij}^0 = E_{ik}^0 = 1$ ,

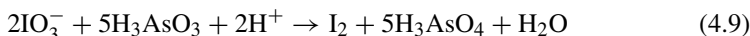
$$C_i = \frac{1}{1 + \frac{1}{2A_j}} + \frac{1}{1 + \frac{1}{2A_k}} = \begin{cases} 4/3 & A_j = A_k = 1 \\ 2/3 & A_j = 1, A_k = 0 \\ 2/3 & A_j = 0, A_k = 1 \\ 0 & A_j = A_k = 0 \end{cases} \quad (4.8)$$

We see that only for  $A_j = A_k = 1$  does  $C_i$  exceed 1; that is, if and only if neurons  $j$  and  $k$  are excited, then and only then is neuron  $i$  excited. This is an AND gate. Other logic gates can be constructed similarly, such as OR, NOR, and so on, by choosing appropriate values of  $K$  and  $E_{ij}^0$  in this simple model.

## 4.2 Implementation of Computers by Macroscopic Chemical Kinetics

With macroscopic chemical kinetics we can construct a Turing machine, a universal computer, which can add two strings of 0 and 1, and hence can subtract, multiply and divide, and carry out complicated calculations, as done by hand computers, laptops, and mainframe machines (see [1–4]).

A different kind of computing device is a parallel machine that can also be implemented by means of macroscopic chemical kinetics. For this purpose we choose a bistable chemical reaction, the iodate–arsenous acid reaction:



When this reaction is run in an open system—a so-called continuous-flow stirred tank reactor, or CSTR (fig. 4.4), with continuous influx of reactants and outflow products and unreacted reactants—then for certain influx conditions the system may be in one of two stationary states far from equilibrium: one of high  $\text{I}_2$  concentration, made visibly blue by addition of some starch, and one of low  $\text{I}_2$  concentration, a colorless solution. Measurements of bistability and chemical hysteresis in this system are shown in fig. 4.5. Bistable reaction systems have some similarities with bistable electronic switches, as pointed out some years ago by Roessler (see cited references in [1–3]). With bistable electronic switches it is possible to build an electronic computer, and now



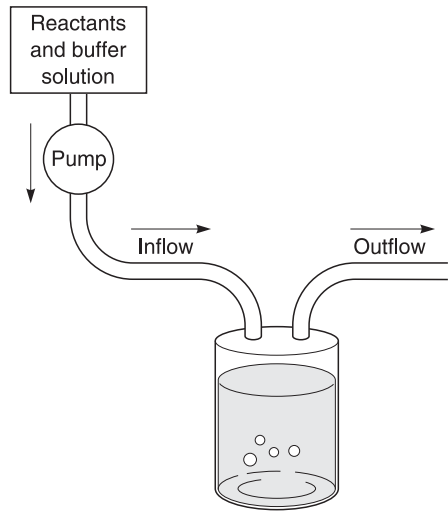


Fig. 4.4 Schematic of a continuous-flow stirred tank reactor (CSTR). The reacting fluid in the beaker is stirred.

we show the construction of a parallel computer, a pattern recognition device, with bistable chemical systems.

Suppose we take 8 CSTRs, each run as shown in fig. 4.6 with the iodate–arsenous acid reaction, eq. (4.8). Each circle is a CSTR containing this bistable reaction. The arrows indicate tube connections among the 8 tanks through which the reaction fluid from one CSTR is pumped at a set rate into another CSTR. The widths of the lines are a qualitative measure of the rate of transport from one CSTR to another. Each isolated reactor can be in one of two stable stationary states; 8 reactors can be in  $2^8$  such states. By our choice of the pumping rates we determine how many stable stationary states there are in the coupled reactor system. The dark (white) circles denote a state of high (low) iodide concentration. The choices of pumping rates and stable stationary states

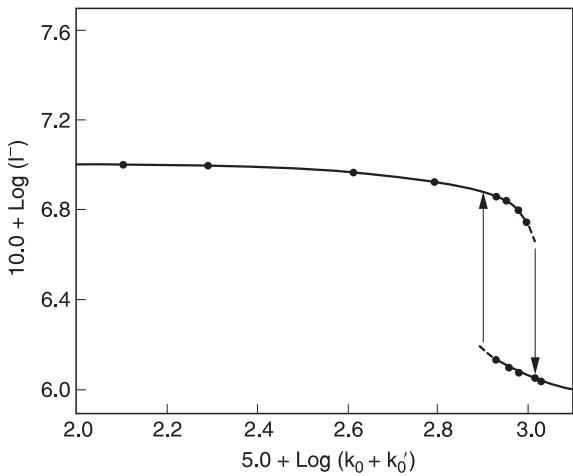


Fig. 4.5 Plot of measured  $(I^-)$  versus inflow rate coefficient  $k_0$  in the iodate–arsenous acid reaction run in an open, well-stirred system, a CSTR. The arrows indicate observed transitions from one branch of stable stationary states to the other stable branch, as the inflow rate coefficient is varied, and define the hysteresis loop. (Taken from [21] with permission.)

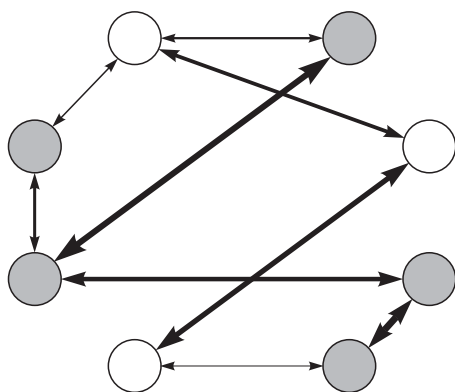


Fig. 4.6 Schematic of 8 CSTRs, each filled with the iodate–arsonous acid reaction, connected by mass flow pumps (arrows); the varying thickness of the arrows denotes varying mass flow rates. Dark circle: the tank is in a high iodide stationary state; white circle: the tank is in a low iodide stationary state.

are made by a Hebbian rule familiar in neural network theory [1,2], which provides stronger connections between reactor tanks of the same color and weaker connections otherwise.

Experiments were made on this system [8] and some results are shown in fig. 4.7. The selected stable stationary states of the system are given in the last column at  $t = 120$  min: the first reactor is blue (marked in black in fig. 4.7), the second colorless (marked in white), and so on. The system was started at time  $t = 0$ , not in one of its stationary states but in one of its stationary states with 2 errors. The errors are in reactors 1 and 2, and the initial state is not one of the stationary states of the coupled system. We are giving the system a pattern of blue and colorless states, with two errors, and we ask the system to correct these errors and identify the correct pattern as one of its stationary

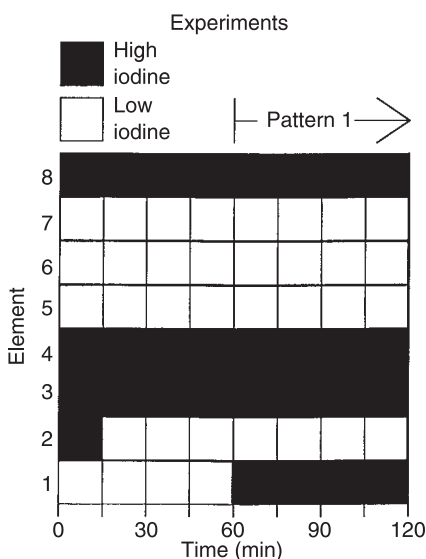


Fig. 4.7 Experiments on pattern recognition. Three patterns are stored in the network. The system is started with one of these patterns but with errors in cells 1 and 2. The correct pattern is recovered after 60 min. (From [8].)

states. As the figure shows, this task is accomplished in time. The correct identification can be achieved for a given number of reactors only if the number of patterns (stable stationary states) given to the system is not too large (about 13% of the total number of reactors) and the number of errors is limited.

This experiment showed the possibility of implementing a parallel computing machine by macroscopic chemical kinetics. In a parallel machine all reactors (the analogs of electronic elements) interact at all times during the computation. In a sequential computer, such as a Turing machine, each element, such as a given logic gate, interacts with only a few other elements at any time, and does not act at all times.

In neural networks, implemented by electronic components, the connections are made by wires, and the connection strengths are determined by resistors. In a network implemented by macroscopic chemical kinetics the connections are made by mass transfer and the connection strengths are pumping rates.

### 4.3 Computational Functions in Biochemical Reaction Systems

If chemical reaction mechanisms running far from equilibrium, such as bistable systems, can perform computational functions, then we may ask whether such computations occur in well-known biochemical reaction mechanisms, and if they do, what role and purpose they serve. Biochemical reaction networks (BRN), such as glycolysis and the tricarboxylic acid cycle, are an integral part of the machinery with which an organism maintains itself and adapts to its environment. These networks are responsible for numerous cellular tasks such as the maintenance of homeostasis and the creation and propagation of chemical signals of many types, such as indications of hunger or satiation.

It is often difficult to determine the basis of the logic of these regulation processes, even for small parts of a BRN. First, the subnetwork may be highly connected with feedback loops and branching pathways. Second, the many kinetic parameters of the BRN are difficult to measure and may be poorly, or not at all, known. Third, the great range of temporal and spatial scales over which a large BRN can react to the perturbations of its variables makes it difficult to deduce biological control and signal processing from models of the system. Hence we seek new methods for the study of the reaction mechanisms of these complex systems, and we proceed with the analysis of some examples of computational functions in BRNs.

Consider the biochemical reaction mechanism shown in fig. 4.8, which occurs in glycolysis [7]; note the similarity in the structure of this mechanism to that of a model of a chemical neuron shown in fig. 4.1. Calculations of the stable stationary states of the biochemical system of fig. 4.8 are plotted in fig. 4.9. The change from high to low concentration of either compound is not as abrupt as that in fig. 4.2 for the chemical model, but is clearly present. We have identified a computational element, a “fuzzy” chemical neuron, in a biochemical reaction.

Logic gates can be constructed by a variety of chemical reaction mechanisms; for example, relatively simple enzymatic reactions can serve as logic gates [7]. Another important unit consists of cyclic enzymatic reactions, as shown in fig. 4.10. This system is present in glycolysis. We single it out since measurements on its stationary states away from equilibrium have been made and compared with calculations [9]. The experiments

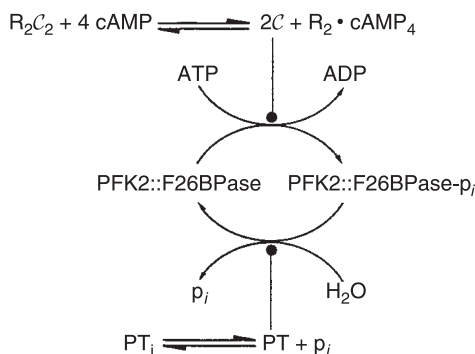


Fig. 4.8 A diagram of the phosphorylation/dephosphorylation cycle of the tandem enzyme PFK2::F26BPase. The lines ending in filled circles indicate that  $C$  and  $PT$  are catalysts for the phosphorylation and dephosphorylation reactions, respectively. This mechanism has a reaction topology very similar to the reaction mechanism in fig. 4.1. Here,  $R_2C_2$  represents the tetrameric enzyme, cAMP-dependent protein kinase.  $R$  denotes the regulatory subunit and  $C$  denotes the catalytic subunit.  $PT$  denotes a generic phosphatase which is inhibited by the presence of an inorganic phosphate. The inactive phosphatase is denoted by  $PT_i$ . The activity of the PFK2 moiety of the tandem enzyme is reduced by phosphorylation whereas that of the F26BPase is increased. This figure is similar to fig. 1 of [22]. (From [7].)

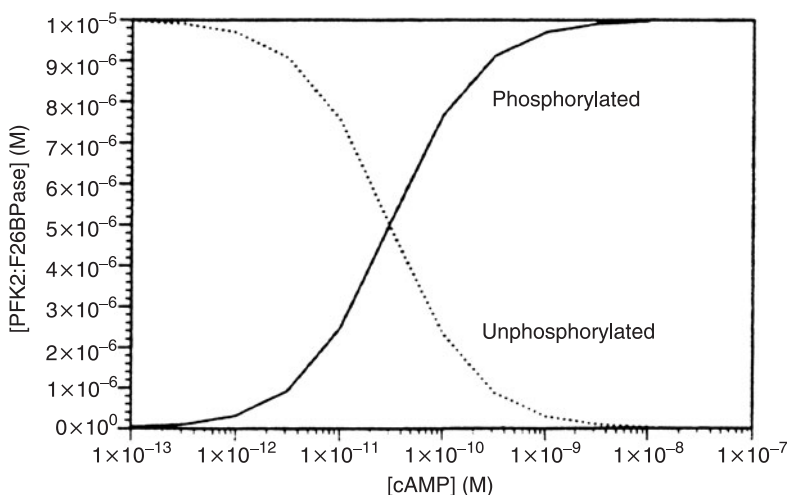


Fig. 4.9 Plot of the steady-state levels of phosphorylated and unphosphorylated PFK2::F26BPase (fig. 4.8) as a function of the concentration of the catalytic subunit of protein kinase  $C$ . The transition between low and high concentrations of phosphorylated protein is gradual, unlike the formal neuron shown in fig. 4.2. The kinetics for this system are taken from [22]. (From [7].)

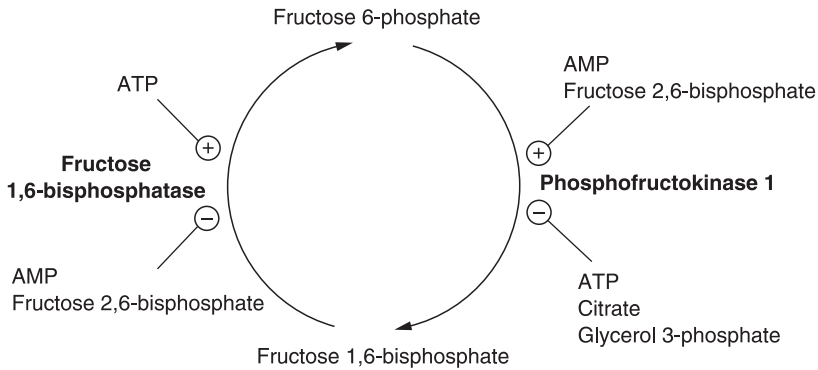
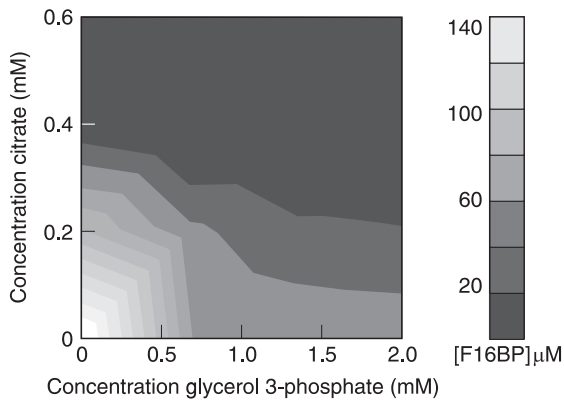


Fig. 4.10 Schematic diagram of the fructose 6-phosphate/fructose 1,6-bisphosphate cycle. Some effectors of the enzymes are noted by lines ending in circles; ⊕ indicates activators and ⊖ indicates inhibitors. Enzymes are noted in bold type. (From [9].)

were made in a CSTR: buffered F6P was flowed into the CSTR and the concentrations of AMP, citrate, F16BP, F26BP, and G3P were measured by means of capillary electrophoresis, at a given stationary state determined by the inflow. The concentrations of the enzymes were kept constant.

An example of these measurements is given in fig. 4.11, which shows the concentration of F16BP, the gray scale on the right and in the square, as a function of the concentration of citrate and G3P. This compares very well with a calculation for this system (fig. 4.12). Both the measurements and the calculation show the presence of a NOR gate. If the concentration of F16BP is small then the gate is off; this occurs for all concentrations of both citrate and G3P unless both of these are small, and then the gate is on and the concentration of F16BP is large. We can correlate the F16BP concentration with the output of this system to be glycolysis or gluconeogenesis. Glycolysis is the conversion of glucose to tricarboxylic molecules and the production of the energetic

Fig. 4.11 Experimentally determined plot of the concentration of F16BP versus the concentrations of effectors citrate and G3P. Black is the lowest concentration of F16BP and white is the highest (see the shaded quantitative scale next to the plot). (From [9].)



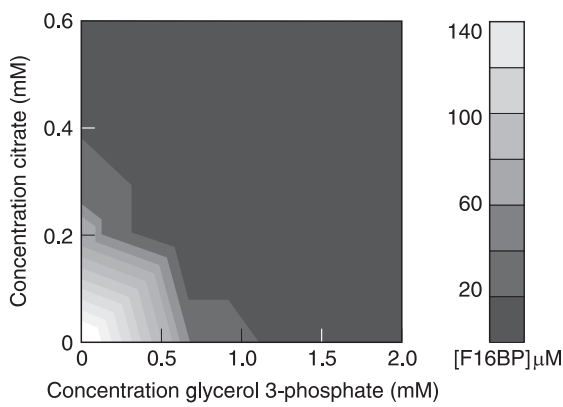


Fig. 4.12 Numerically determined plot of the concentration of F16BP versus the concentrations of effectors citrate and G3P. Black is the lowest concentration of F16BP and white is the highest. (From [9].)

species ATP from ADP and AMP. This occurs for high concentrations of F16BP. Gluconeogenesis is a reverse process of glycolysis and occurs for high concentrations of F6P, that is, low concentrations of F16BP.

If we think of a formal combination of two structures as shown in fig. 4.10 [7], we come to an important part of glycolysis, the hexose phophate system (fig. 4.13). The stationary states of this system (fig. 4.14) constitute a fuzzy aggregation function: an OR gate at low concentrations of citrate and cAMP, and an AND gate at high concentrations. This biological reaction mechanism is a control point in glycolysis/gluconeogenesis. The enzymes in this kinetic mechanism are under allosteric control of many of the

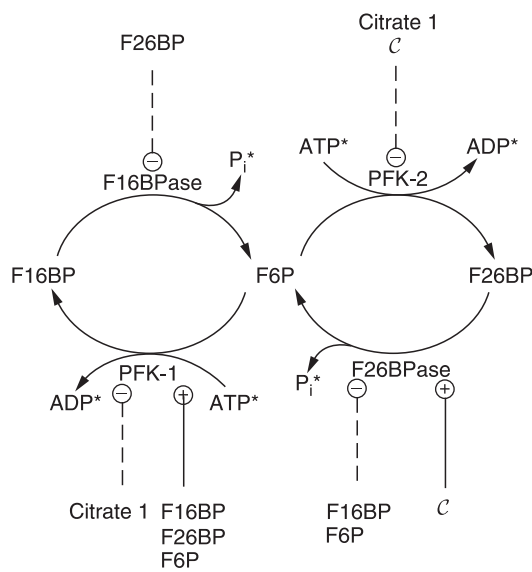


Fig. 4.13 Interconversion of fructose 6-phosphate with fructose 1,6-bisphosphate and fructose 2,6-bisphosphate. Species marked by asterisks are considered to be held at constant concentration by buffering or flow. Positive (negative) effectors are connected to their enzymes by a line ending with a circle in which there is a plus (minus) sign. (From [7].)

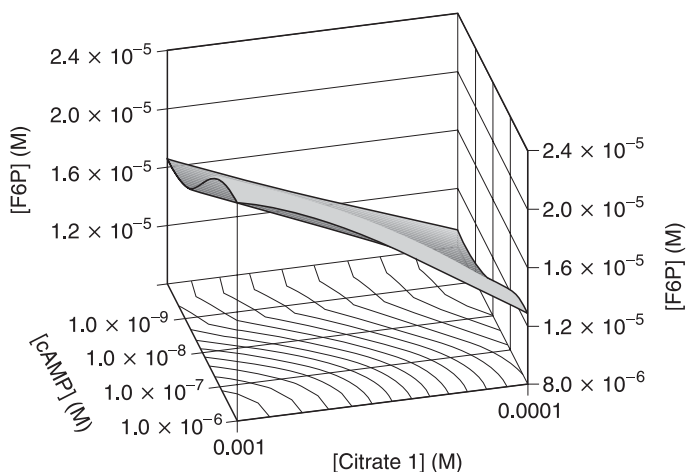


Fig. 4.14 Plot of the results of a calculation of the steady-state concentration of fructose 6-phosphate for the system shown in fig. 4.13. The enzyme models are either based on Michaelis–Menten formalisms or modifications of multiple allosteric effector equations. The gate exhibits a function with both AND and OR properties. At low concentrations of both inputs, the mechanism functions similarly to an OR gate, while at simultaneously high concentrations of the input species (citrate and cAMP), the output behavior more closely resembles a fuzzy logic AND gate. The mechanism satisfies the requirements for a fuzzy aggregation function. (From [7].)

chemical signals of cellular energy states, such as the adenosine phosphates, cAMP, citrate, and NAD.

For another implementation of electronic devices by means of macroscopic kinetics we mention the construction of frequency filters and signal processing [14], but do not explore this subject further.

We have seen that computations can be achieved by chemical and biochemical reaction mechanisms, and have located computational functions in biological reaction systems. This identification helps in understanding functions and control in such systems. It also helps in suggesting new approaches to the determination of causal connectivities of reacting species, of reaction pathways, and reaction mechanisms, by exploring analogs of investigations in electronics, system analysis [15–17], multivariate statistics [18–20], and other related disciplines. We begin with a study of the response of chemical kinetic systems to pulse perturbations of the concentration of one or more chemical species.

## References

- [1] Hjelmfelt, A.; Weinberger, E. D.; Ross, J. Chemical implementation of neural networks and Turing machines. *Proc. Natl. Acad. Sci. USA* **1991**, 88, 10983–10987.
- [2] Hjelmfelt, A.; Weinberger, E. D.; Ross, J. Chemical implementation of finite-state machines. *Proc. Natl. Acad. Sci. USA* **1992**, 89, 383–387.

- [3] Hjelmfelt, A.; Ross, J. Chemical implementation and thermodynamics of collective neural networks. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 388–391.
- [4] Hjelmfelt, A.; Schneider, F. W.; Ross, J. Pattern recognition in coupled chemical kinetic systems. *Science* **1993**, *260*, 335–337.
- [5] Hjelmfelt, A.; Ross, J. Implementation of logic functions and computations by chemical kinetics. *J. Phys. D.* **1995**, *84*, 180–193.
- [6] Hjelmfelt, A.; Ross, J. Pattern recognition, chaos, and multiplicity in neural networks of excitable systems. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 63–67.
- [7] Arkin, A.; Ross, J. Computational functions in biochemical reaction networks. *Biophys. J.* **1994**, *67*, 560–578.
- [8] Laplante, J. P.; Pemberton, M.; Hjelmfelt, A.; Ross, J. Experiments on pattern recognition by chemical kinetics. *J. Phys. Chem.* **1995**, *99*, 10063–10065.
- [9] Hauri, D. C.; Shen, P.; Arkin, A. P.; Ross, J. Steady-state measurements on fructose 6-phosphate/fructose 1,6-bisphosphate interconversion cycle. *J. Phys. Chem. B* **1997**, *101*, 3872–3876.
- [10] Okamoto, M.; Sakai, T.; Hayashi, K. Switching mechanism of a cyclic enzyme system: role as a “chemical diode.” *BioSystems* **1987**, *21*, 1–11.
- [11] Okamoto, M.; Hayashi, K. Dynamic behavior of cyclic enzyme systems. *J. Theor. Biol.* **1983**, *104*, 591–598.
- [12] Okamoto, M.; Katsurayama, A.; Tsukiji, M.; Aso, Y.; Hayashi, K. Dynamic behavior of system realizing two-factor model. *J. Theor. Biol.* **1980**, *83*, 1–16.
- [13] Okamoto, M.; Hayashi, K. Control mechanism for a bacterial sugar-transport system: theoretical hypothesis. *J. Theor. Biol.* **1985**, *113*, 785–790.
- [14] Samoilov, M.; Arkin, A.; Ross, J. Signal processing by simple chemical systems. *J. Phys. Chem. A* **2002**, *106*, 10205–10221.
- [15] Klir, G. J. Reconstructability analysis: overview and bibliography. *Int. J. Gen. Sys.* **1981**, *7*, 1–6.
- [16] Conant, R. C. Extended dependency analysis of large systems, Part II: Static analysis. *Int. J. Gen. Sys.* **1988**, *14*, 125–141.
- [17] Conant, R. C. Extended dependency analysis of large systems, Part I: Dynamic analysis. *Int. J. Gen. Sys.* **1988**, *14*, 97–123.
- [18] Marriot, F. H. C. *The Interpretation of Multiple Observations*; Academic Press: New York, 1974.
- [19] Mardia, K. V.; Kent, J. T.; Bibby, J. M. *Multivariate Analysis*; Academic Press: San Francisco, 1979.
- [20] Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes in C*; Cambridge University Press: New York, 1988.
- [21] Ganapathisubramanian, N.; Showalter, K. Bistability, mushrooms, and isolas. *J. Chem. Phys.* **1984**, *80*, 4177–4184.
- [22] Schacter, E.; Chock, P. B.; Stadtman, E. R. Regulation through phosphorylation/dephosphorylation in cascade systems. *J. Biol. Chem.* **1984**, *259*, 12252–12259.



# 5

## Response of Systems to Pulse Perturbations

---

### 5.1 Theory

Consider a chemical reaction system with many chemical species; it may be in a transient state but it is easier to think of it in a stable stationary state, not necessarily but usually away from equilibrium. We wish to probe the responses of the concentrations of the chemical species to a pulse perturbation of one of the chemical species [1]. The pulse need not be small; it can be of arbitrary magnitude. This is analogous to providing a given input to one variable of an electronic system and measuring the outputs of the other variables. The method presented in this chapter gives causal connectivities of one reacting species with another as well as regulatory features of a reaction network. Much more will be said about the responses of chemical and other systems to pulses and other perturbations in chapter 12. The effects of small perturbations on reacting systems have been investigated in a number of studies [2–5], to which we return in chapters 9 and 13.

Let us begin simply: Consider a series of first-order reactions as in fig. 5.1, which shows an unbranched chain of reversible reactions. We shall not be restricted to first-order reactions but can learn a lot from this example. Let there be an influx of  $k_0$  molecules of  $X_1$  and an outflow of  $k_8$  molecules of  $X_8$  per unit time. We assume that the reaction proceeds from left to right and hence the Gibbs free energy change for each step and for the overall reaction in that direction is negative. The mass action law for the kinetic equations, say that of  $X_2$ , is

$$\frac{dX_2}{dt} = k_1X_1 + k_{-2}X_3 - (k_{-1} + k_2)X_2 \quad (5.1)$$

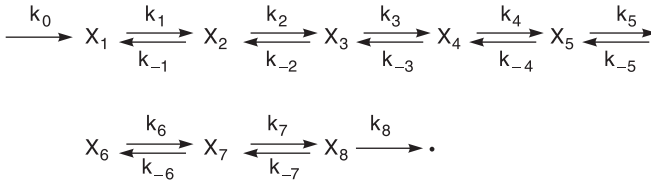


Fig. 5.1 An unbranched chain of reversible first-order reactions. (From [1].)

If all the time derivatives of the concentrations are zero, then the system is in a stationary state. Suppose we perturb that stationary state with an increase in  $X_1$  by an arbitrary amount and solve the kinetic equations numerically for the variations of the concentrations as a function of time, as the system returns to the stationary state. A plot of such a relaxation is shown in fig. 5.2. To find in that relaxation the extremum value of  $X_2$ , denoted by  $X_2^*$ , we set the time derivative in eq. (5.1) equal to zero and have

$$k_1 X_1 + k_{-2} X_3 = (k_{-1} + k_2) X_2^* \quad (5.2)$$

The values of  $X_1$  and  $X_3$  are those at the extremum of  $X_2$ , that is, at  $X_2^*$ .

Next, suppose that the system is in a stable stationary state, but not at equilibrium. The time derivative of each concentration  $X$  is zero, and there is a constant flux of mass into and out of the system. The concentrations at a stationary state are denoted by  $X^s$ .

It is advantageous to work with relative changes of concentrations defined by

$$u = \frac{(X - X^s)}{X^s} \quad (5.3)$$

(There is a misprint in the definition of  $u$  in [1].)

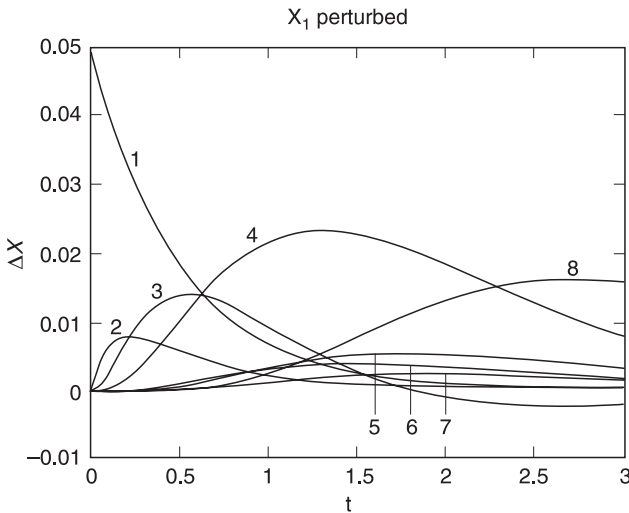


Fig. 5.2 Plot of absolute deviation of the concentrations for the species in the sequence of reactions in fig. 5.1 as a function of time after a pulse of  $X_1$  at  $t = 0$ .

Substitution of eq. (5.3) into eq. (5.2) for the extremum value of  $X_2$  yields

$$(k_{-1} + k_2)(u_2^* X_2^s + X_2^s) = k_1(u_1 X_1^s + X_1^s) + k_{-2}(u_3 X_3^s + X_3^s) \quad (5.4)$$

At the stationary state all time derivatives are zero and hence from the equation for  $X_2$  we have

$$k_1 X_1^s + k_{-2} X_3^s - (k_{-1} + k_2) X_2^s = 0 \quad (5.5)$$

We subtract eq. (5.5) from eq. (5.4) to obtain

$$(k_{-1} + k_2) X_2^s u_2^* = k_1 X_1^s u_1 + k_{-2} X_3^s u_3 \quad (5.6)$$

Next we introduce the definition

$$j_f = k_1 X_1 \quad (5.7)$$

for the forward flow into  $X_2$ , and

$$j_T = j_f + k_{-2} X_3$$

for the total flow into  $X_2$ , which brings us to the value of the relative concentration of  $X_2$  at the maximum of the response to a pulse perturbation:

$$u_2^* = \alpha u_1(t) + (1 - \alpha) u_3(t), \quad \alpha \equiv j_i / j_T \quad (5.8)$$

A plot of relaxation of the relative concentrations of the species in the system in fig. 5.1 is shown in fig. 5.3(a) after a pulse of species  $X_1$ . There are regularities to be observed in the plot of the concentrations (fig. 5.2), but even more so in the plot of the relative concentration (fig. 5.3):

1. The time of occurrence of an extremum in a given species increases as the number of reaction steps separating that species from the initially perturbed species increases, unless some species act as catalysts, or effectors on catalysts, in distant reactions.

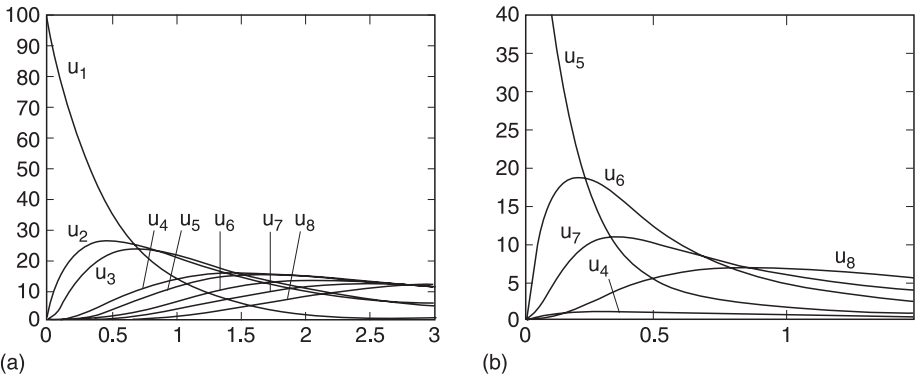


Fig. 5.3 Plots of the relative deviation in concentration from the stationary state versus time for all the species of the mechanism in fig. 5.1. The maxima are ordered according to the number of reaction steps separating that species from the initially perturbed species. In (b), a pulse of  $X_5$  propagates with large amplitude in the direction of the overall reaction velocity and with small amplitude in the opposite direction. (From [1].)

2. The initial response of the relative concentration of a species in time decreases as the number of reaction steps, as described in point 1, increases.
3. Species directly (not directly) connected by reaction to the initially perturbed species have nonzero (zero) slope.
4. All responses of species in the relaxation back to the stationary state are positive deviations from that state unless there is feedback, or feedforward, or there are higher order ( $> 1$ ) kinetic steps.
5. For short times after a pulse, prior to the exit of the material from the pulse, the concentration change of the pulse is conserved; the sum of deviations of concentrations, weighted by stoichiometric coefficients, must be constant and equal to the change in concentration due to the initial pulse. This property is useful in confirming that all species produced from the pulse due to reactions have been detected, and in determining the correct stoichiometric coefficients of reactants and products.
6. When it is possible to identify rate expressions for reaction steps, then rate coefficients may be estimated from the concentration measurements.
7. Equation (5.8) shows that a maximum in  $u_2$  occurs between the relaxation curves of the preceding and succeeding species.
8. A pulse applied to a species in the middle of an unbranched chain, such as species 5 in fig. 5.3(b), propagates well downhill in Gibbs free energy to species 6, 7, and 8 but weekly uphill to species 4 and not past that (for the particular set of rate coefficients chosen for this example).

We are beginning to see from figs. 5.2 and 5.3 and the deductions (points 1–8) how the causal connectivity, at least in this simplest of reaction mechanisms, can be deduced from pulse perturbations.

Further regularities appear for the case of irreversible reactions. Suppose we have the first-order reactions



with the rates for  $X$  given by

$$\frac{dX}{dt} = k_1 X_1 - k_2 X \quad (5.10)$$

At the maximum of the relaxation of  $X$ , labeled  $X^*$ , we have

$$k_1 X_1 = k_2 X^* \quad (5.11)$$

and at the stationary state

$$k_1 X_1^s = k_2 X^s \quad (5.12)$$

Division of eq. (5.10) by eq. (5.12) and addition of 1 to both sides (see eq. (5.3)) yields

$$u_1 = u^* \quad (5.13)$$

The relative concentration of  $X$  at its maximum value in the relaxation to the stationary state equals the concentration of  $X_1$  at that instant of time. This is illustrated for the system in fig. 5.1 for irreversible reactions, that is, the case for which all the reverse rate coefficients are set to zero (see fig. 5.4).

Of course most reactions are not simple first order, but it is important to note that for appropriate isotope experiments the responses to pulses of isotopes are always first order, regardless of the order of the kinetics, second or higher, that is, regardless of the

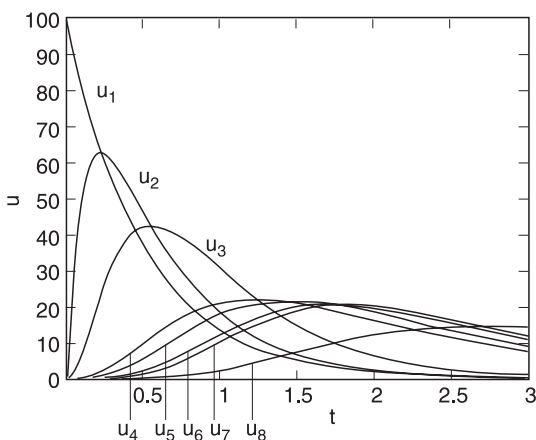


Fig. 5.4 Plot of the relative deviation of concentration in the sequence of reactions in fig. 5.1 for the case of irreversible reactions. (From [1].)

nonlinearities of the kinetics. To show this, we study the kinetics of a species  $A$ ; the system is in a stationary state, equilibrium or not, and we apply a pulse of radioactive  $A$ , labeled  $A^*$ , in such a way that the concentration of the sum of  $A$  and  $A^*$  is constant. We also require that the kinetic isotope effect of this radioactive tracer be negligible, that is, the rate coefficients for the reactions of  $A$  and  $A^*$  are essentially the same.

We assume the rate of disappearance of  $A$  to be given by

$$R = kA^\nu \quad (5.14)$$

We write the stoichiometric equation  $XA^* + (\nu - x)A = \text{products}$ ,  $x = 1, 2, \dots, \nu$ . Hence the mass action law kinetic equation for the time variation of the isotope  $A^*$ , the rate  $R^*$ , is for the case  $\nu = 2$ :

$$\begin{aligned} R^* &= kAA^* + k(A^*)^2 = kA(A + A^*) \\ R^* &= \frac{A^*}{A + A^*} k(A + A^*)^2 = \frac{A^*}{A + A^*} R(A + A^*) \end{aligned} \quad (5.15)$$

Since the sum of the concentrations of  $A$  and  $A^*$  are constant, we see that the rate of change of  $A^*$ , that is,  $R^*$ , is linear, first order, in  $A^*$  regardless of the value of  $\nu$ . This is an important and widely applicable result [6].

Many reaction sequences consist of converging and diverging chains; an example of converging irreversible reactions is shown in fig. 5.5. Calculations of the response of this system to a pulse perturbation of  $X_1$  and  $X_8$  are plotted in figs. 5.6(a) and (b), respectively. The plots indicate the convergence of two chains at the species  $X_3$ .

We proceed with the consideration of a linear chain of coupled first- and second-order reactions (fig. 5.7). If species  $X_1$  is pulsed, then the relaxation of the various species is shown in fig. 5.8. There are interesting approximate relations for such systems among the amplitudes of changes of relative concentrations. Consider the variation of  $X_2$  upon a pulse of  $X_1$  administered to the system; the deterministic rate equations are

$$\frac{dX_2}{dt} = k_1X_1 - k_2X_2^2 \quad (5.16)$$



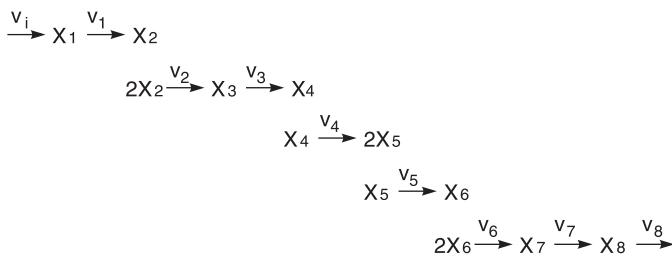
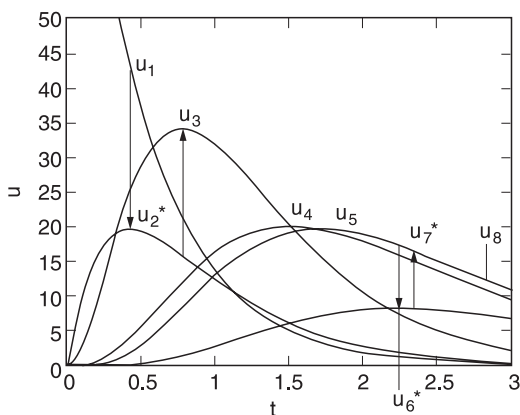


Fig. 5.7 Chemical reaction mechanism of a linear chain of coupled first- and second-order reactions. The order of a reaction is given by the stoichiometry. (From [1].)

Fig. 5.8 Plots of relative deviation in concentration versus time for species of the mechanism in fig. 5.7. A pulse perturbation of the concentration of species  $X_1$  results in the responses shown. The pulse propagates through the chain with maxima of relative deviations of species ordered according to the positions of species in the chain. From the plot we find the relations:  $u_2^* \approx (1/2)u_1$ ,  $u_3^* \approx 2u_2$ ,  $u_6^* \approx (1/2)u_5$ , and  $u_7^* \approx 2u_6$ ; the values of these coefficients may be related to the orders of the reactions through eq. (5.18). (From [1].)



eq. (5.3) we have

$$u_1 = 2u_2^* \quad (5.18)$$

This approximate relation, and others, can also be seen in fig. 5.8, and it yields the stoichiometric coefficient 2 of the reaction  $2X_2 \rightarrow X_3$ . Thus pulse methods can provide information on the stoichiometry of elementary reactions in a reaction mechanism.

Suppose we have a reaction mechanism with feedback, as shown in fig. 5.9. Here species  $X_7$  acts as a catalyst on the reaction  $X_3 \rightarrow X_4$ . If we perturb species  $X_1$ , then the relaxation of all species is shown in fig. 5.10(a), in a plot of concentration deviations  $\Delta X$  as a function of time; at longer times  $\Delta X_3$  becomes negative, which indicates the possibility of feedback. This possibility is confirmed in fig. 5.10(b) in which the catalytic species  $X_7$  is perturbed. The effect of the catalysis makes  $\Delta X_3$  negative and  $\Delta X_4$  positive at once, thus confirming the catalytic effect of species  $X_7$  on this reaction.

Other geometries of reaction pathway have been investigated: branched chains, that is, several reaction chains converging or diverging at some species, and cycles of reactions.

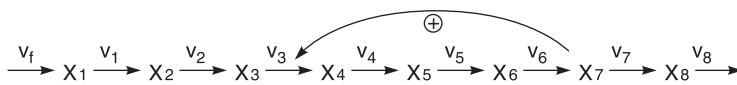


Fig. 5.9 Chemical reaction mechanism illustrating positive feedback in a linear chain of irreversible first-order reactions. The rate of production of species  $X_1$  is held constant at  $v_f$ . (From [1].)

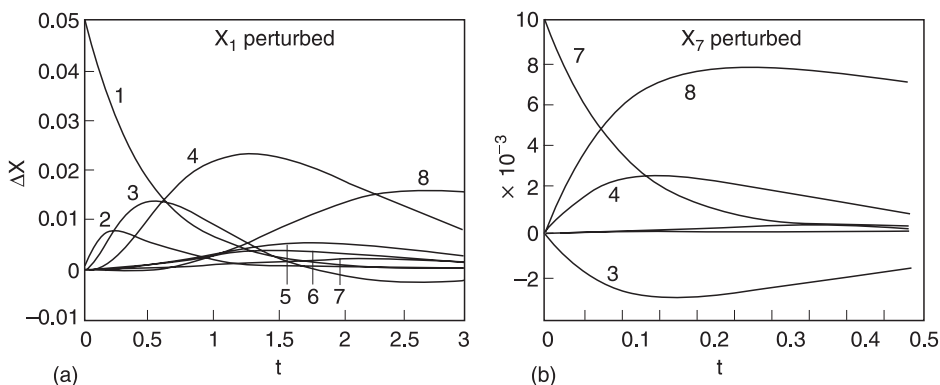


Fig. 5.10 Plots of deviation in concentration  $\Delta X$  from the stationary state value versus time for the species of the mechanism in fig. 5.9. In (a), a perturbation of the concentration of species  $X_1$  causes a pulse to propagate through the linear chain  $X_2, X_3, \dots, X_8$ ; the deviation of species  $X_3$  falls below the stationary state value, which indicates a possible feedback effect from one of the species farther in the chain. In (b), a perturbation of species  $X_7$  causes the initial slopes of the deviations in  $X_3, X_4$ , and  $X_8$  to be nonzero and the deviations in  $X_3$  and  $X_4$  to be mirror images about the time axis (for short times); these observations indicate that  $X_7$  activates the reaction from  $X_3$  to  $X_4$  (because the peak of  $X_4$  occurs before that of  $X_3$ , implying that  $X_3$  precedes  $X_4$  in the reaction sequence) and is the precursor to  $X_8$ . The maximum deviations are approximately 10% of the stationary state values or larger, except  $X_4$  in (b), which is 2% (the stationary state values of  $X_1$ – $X_8$  are 0.05, 0.0125, 0.0166, 0.1, 0.0166, 0.025, 0.01, and 0.1, respectively). (From [1].)

## 5.2 An Example: The Glycolytic Pathway

An application of the pulse method outlined above is instructive. We choose as a model a part of the glycolytic pathway shown in fig. 5.11. This model includes many of the known activations and inhibitions of enzymes by metabolites. The hexokinase reaction is assumed to be at a constant rate, which neglects the influence of ATP and glucose 6-phosphate. The concentrations of glucose, lactic acid, and the adenine nucleotide pool (the sum of the concentrations of AMP, ADP, and ATP) are taken to be constant. We represent the kinetics of this system by deterministic rate equations (see [1] for details), apply a pulse of concentration of one species to the system in a nonequilibrium stationary state, and calculate the response of the following independent species: fructose



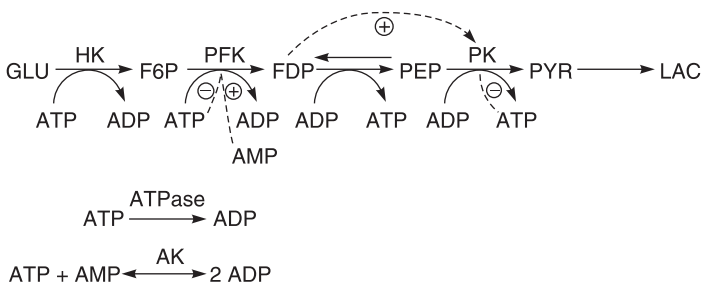


Fig. 5.11 Chemical reaction mechanism for an abbreviated model of the glycolytic pathway. The mechanism includes many of the known activations and inhibitions of enzymes by metabolites. Broken lines indicate activation ( $\oplus$ ) or inhibition ( $\ominus$ ) of enzymes by metabolites. The abbreviations for the enzymes are hexokinase (HK), phosphofructokinase (PFK), and pyruvate kinase (PK); the abbreviations for the five independent variables are fructose 6-phosphate (F6P), fructose 1,6-bisphosphate (FDP), phosphoenolpyruvate (PEP), pyruvate (PYR), and ATP (in fig. 5.12, these are labeled 1–5, respectively). (From [1].)

6-phosphate (F6P, labeled 1); fructose 1,6-bisphosphate (FDP, 2); phosphoenolpyruvate (PEP, 3); pyruvate (PYR, 4); and adenosine triphosphate (ATP, 5).

In fig. 5.12 we show the time series for the relaxation of the absolute deviation in concentration of all the five species following a pulse perturbation of one species at a time. In part (a) of that figure is plotted the response to a perturbation of species 1. Species 2, 3, and 4 have maximum response in time in succession, which tells us that species 2 follows 1, 3 follows 2, and 4 follows 3. Species 3 and 4 have zero initial slope consistent with this assignment. The deviation in 2 starts at zero, which implies that the rate of loss of 2 into any species, that is 3 and 4, is initially zero. The deviations in 3 and 4 are mirror images, which indicates that the conversion of 3 to 4 is affected by 2. If species 1 were to affect the conversion of 3 to 4, then the initial slope of 3 and 4 would be nonzero. A verification of the conversion of 1 to 4 in steps is shown in fig. 5.13; the sum of the deviations in concentrations,  $\Delta X_1 + \Delta X_2 + (\Delta X_3 + \Delta X_4)/2$ , is nearly constant, decaying slowly in time. Thus there is an approximate conservation of mass among these species, in which each molecule of 2 produces two molecules of 3, and each molecule of 3 produces one molecule of 4. The initial slope of species 5 is equal to the negative of the initial slope of 2, from which we deduce that 5 is consumed in equal numbers to 2 produced.

The responses of all species to a pulse perturbation of species 2 are shown in fig. 5.12(b). The initial slopes of 3 and 4 are nonzero, which indicates that species 2 has a direct effect (activation) on the conversion of 3 to 4, and 2 slowly converts to 3. The minimum in 3 precedes the maximum in 4; 2 directly produces 3. If that were not the case, then the minimum in 3 would follow the maximum in 4, which is not the case. Species 1 does not respond to this perturbation, which shows that the reaction of 1 to 2 is highly irreversible. The initial slope of 5 equals that of 4, and hence 5 is produced at the same rate as 4 in the conversion of 3 to 4, which is new information.

Figure 5.12(c) shows the responses of all the species to a pulse perturbation of species 3. Both species 2 and 4 have nonzero initial slopes and from that we deduce

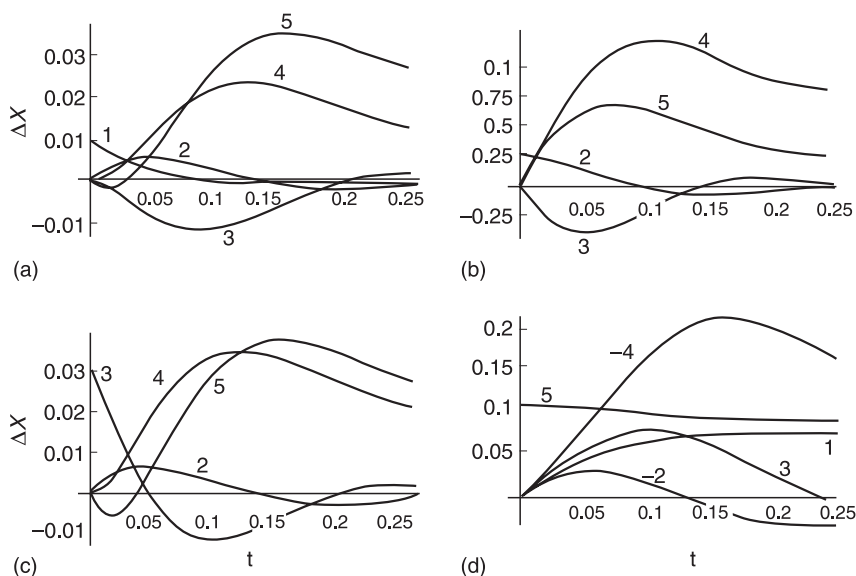


Fig. 5.12 Plots of time series for the absolute deviations in concentrations of species of the mechanism in fig. 5.11 with impulse perturbations of (a) F6P (1), (b) FDP (2), (c) PEP (3) and (d) ATP (5). From an analysis of the responses to these perturbations, we deduce: 1 produces 2, 2 produces 3, 2 activates the conversion of 3 to 4, 3 produces 4, and 5 is consumed in the conversion of 1 to 2 (with the stoichiometric ratio of species 5 to species 2 being 1:1); the conversion of 1 to 2 is highly irreversible; 5 is produced at the same rate as 4 in the reaction 3 to 4; 5 is consumed at the same rate as 2 is produced, which indicates that 5 is a cosubstrate for the reverse reaction of 3 to 2; 5 inhibits the reaction of 1 to 2 and the reaction of 3 to 4. In each of the response plots, the perturbation is 10% of the stationary state value; responses are approximately several percent of the stationary state values or larger, except  $X_5$ , which is in (a) and (b) (the stationary state values of  $X_1$ – $X_5$  are 0.091, 0.200, 2.70, 3.00, and 30.0, respectively). (From [1].)

that 3 produces 2 and 4. The small but present amplitude of 2 shows that the reaction 2 to 3 is reversible. Species 5 is consumed at the same rate as species 2 is produced, and thus 5 is a cosubstrate for the reaction 3 to 2. There is another transient conservation of mass in species 2, 3, and 4 (fig. 5.13b); the initial flow is from 3 to  $(1/2)2$ , and over a longer time from 3 to 4.

Finally, in fig. 5.12(d) we plot the responses of all species to a pulse perturbation of species 5. The minus signs in the figure denote reflection of deviations about the time axis. The increase in 1 and the decrease in 2 show that 5 inhibits the reaction 1 to 2, without affecting the rate of 2 to 3. The opposite rates of 3 and 4 indicate that 5 also inhibits the reaction of 3 to 4.

A summary of all the information obtained from the analysis of the results of the calculations of the responses to pulse experiments on this model is given in fig. 5.14. There is a chain from species 1 to 2 to 3 to 4. ATP is a substrate in the formation of 1 and 2; it is a product in the reaction of 2 to 3, and 3 to 4; and it is an inhibitor of the reactions of 1 to 2, and 3 to 4. The formation of 1 is irreversible, and so are the reactions

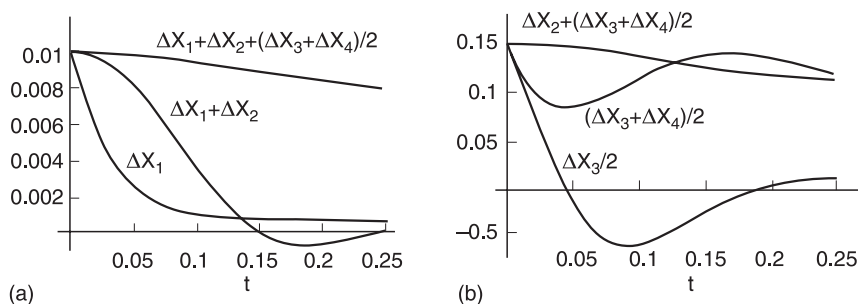


Fig. 5.13 Plots of time series for sums of deviations in concentrations for the species of the mechanism in fig. 5.11. In (a), the concentration of species  $X_1$  is perturbed; in (b), the concentration of species  $X_3$  is perturbed. Plot (a) shows that the sum of the deviations,  $\Delta X_1 + X_2 + (\Delta X_3 + \Delta X_4)/2$ , is slowly decaying in time. From this conservation, we infer that species 2 produces two molecules of species 3, which produces species 4. Plot (b) shows a similar transient conservation of mass in species 2, 3, and 4 following a perturbation of 3; the initial flow is from 3 to  $(1/2)2$  and over a longer time from 3 to 4. (From [1].)

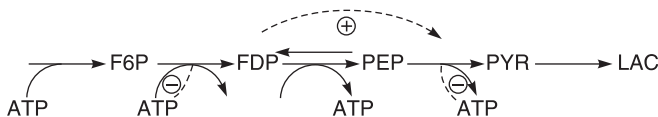


Fig. 5.14 A simplified mechanism constructed from the analysis of figs. 5.12 and 5.13. This diagram captures the topology of the reaction network in fig. 5.11 and many of the effectors. (From [1].)

1 to 2, 3 to 4, and 4 to 5. The reaction 2 to 3 is reversible. Species 2 is an activator of the reaction 3 to 4.

We began with the model given in fig. 5.11, applied pulse perturbations, and from the results constructed fig. 5.14. A comparison of these two figures shows that the results of the pulse perturbations capture the topology and the causal connectivities of the reaction network and many of the effectors, both positive and negative. Further studies of adenine nucleotides are necessary to determine the roles of AMP and ADP.

**Acknowledgment** This chapter is based on selections from the article “Determination of causal connectivities of species in reaction networks” by William Vance, Adam Arkin, and John Ross [1], with some changes in wording.

## References

- [1] Vance, W.; Arkin, A.; Ross, J. Determination of causal connectivities of species in reaction networks. *Proc. Natl. Acad. Sci. USA* **2002**, 99, 5816–5821.
- [2] Tyson, J. J. Classification of instabilities in chemical reaction systems. *J. Chem. Phys.* **1975**, 62, 1010–1015.

- [3] Luo, Y.; Epstein, I. R. Feedback analysis of mechanisms for chemical oscillators. *Adv. Chem. Phys.* **1990**, 79, 269–299.
- [4] Chevalier, T.; Schreiber, I.; Ross, J. Toward a systematic determination of complex reaction mechanisms. *J. Phys. Chem.* **1993**, 97, 6776–6787.
- [5] Mihaliuk, E.; Skodt, H.; Hynne, F.; Sorenson, P. G.; Showalter, K. Normal modes for chemical reactions from time series analysis. *J. Phys. Chem.* **1999**, 103, 8246–8251.
- [6] Neiman, M. B.; Gál, D. *The Kinetic Isotope Method and Its Application*; Akadémic Kiadó: Budapest, 1971.

## Experimental Test of the Pulse Perturbation Method for Determining Causal Connectivities of Chemical Species in a Reaction Network

---

For an experimental test of the pulse perturbation method [1] we choose a part of glycolysis shown in fig. 6.1. There are similarities and some differences between the model in fig. 5.12 and the reaction system in fig. 6.1. The reaction system has reactants, enzymes, and some effectors. One point of interest in choosing this system is the test of detecting and identifying the split of the reaction chain, from glucose to F1,6BP, at the aldolase reaction into two chains, one terminating at G3P and the other at 3PG.

The experiments were run in a continuous-flow stirred tank reactor (CSTR) (fig. 6.2) with the reaction system at a nonequilibrium stationary state, such that the reactions run spontaneously from glucose to G3P and 3PG. The concentrations of the species at this state are close to those of physiological conditions. The metabolites G6P, F6P, F1,6BP, DHAP, G3P, and 3PG were detected and analyzed by capillary electrophoresis. Typical relative errors were 4% for G6P, 11% for F6P, 15% for F1,6BP, 9% for DHAP, 6% for 3PG, and 3% for G3P.

Figure 6.3 shows the responses of the species to a pulse of G6P, in a plot of relative concentrations versus. time during the relaxation, after the pulse, back to the stationary state. Complete relaxation took about half an hour. As seen from the amplitudes of the responses in the plot, the temporal order of propagation of the pulse is: G6P, F6P, DHAP, G3P, and 3PG. The time ordering of the maximum deviations agrees with this ordering except perhaps for G3P and 3PG. In some experiments, as in this one, the species F1,6BP could not be measured adequately and is not shown. It is possible to extract qualitative information on rates but difficult to derive quantitative information.

Following a pulse of F1,6BP (fig. 6.4), the temporal order of propagation in the maximum relative concentrations is F1,6BP, DHAP, and with similar amplitudes G6P

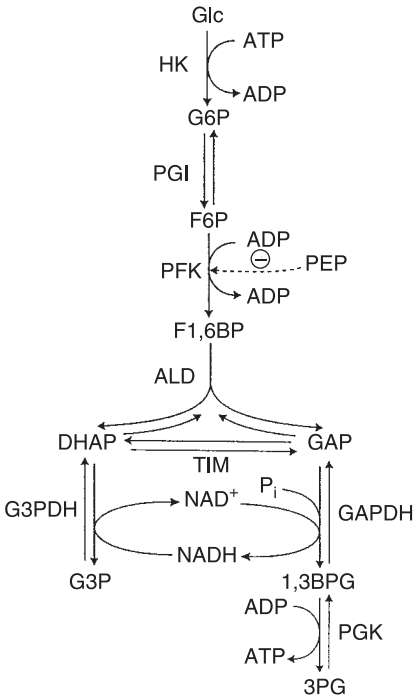


Fig. 6.1 Reaction scheme of the system under study. Glc, glucose; HK, hexokinase; PGI, phosphoglucose isomerase; PFK, phosphofructokinase; ALD, aldolase; TIM, triose phosphate isomerase; G3PDH, glycerol 3-phosphate dehydrogenase; 1,3BPG, 1,3-bisphosphoglycerate; PGK, phosphoglycerate kinase; G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; ATP, adenosine triphosphate; ADP, adenosine diphosphate; PEP, phosphoenolpyruvate; F1,6BP, fructose 1,6-bisphosphate; DHAP, dihydroxyacetone phosphate; G3P, glycerol 3-phosphate; GAP, glyceraldehyde 3-phosphate; 1,3BPG, 1,3-bisphosphoglycerate; 3PG, 3-phosphoglycerate; NADH, nicotinamide adenine dinucleotide (reduced form);  $\text{NAD}^+$  (oxidized form). (From [1].)

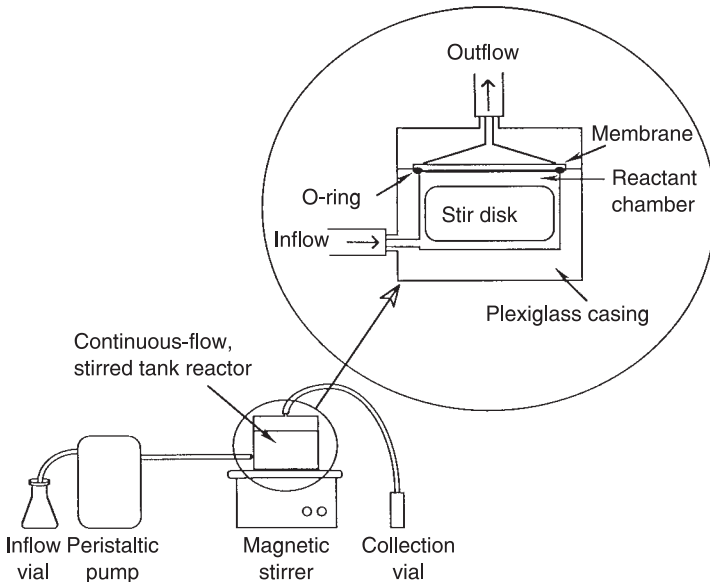


Fig. 6.2 Schematic drawing of a continuous-flow stirred tank reactor (CSTR).

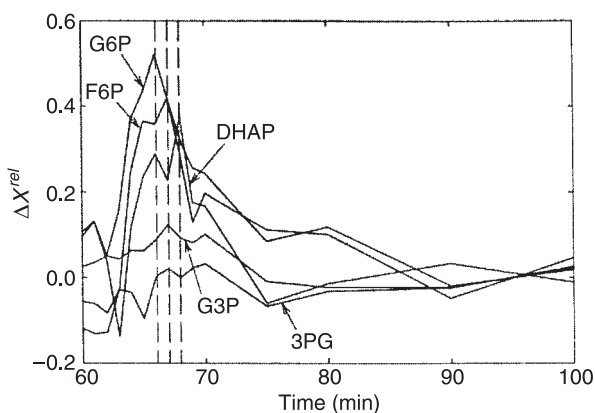


Fig. 6.3 Relative variations of several concentrations after a pulse of G6P. (From [1].)

(slightly higher), G3P, 3PG, and F6P (slightly lower). These small differences were within errors of measurement and are therefore not significant. In this experiment the measurements of F1,6BP are reliable. The intercepts of successive curves, in order of amplitudes, support the same propagation order as that of the maximum relative concentration: the relaxation curve of DHAP crosses that of F1,6BP before G3P and 3PG intercept DHAP.

The responses to a pulse of DHAP are plotted in fig. 6.5. The maximum response is that of F1,6BP. The maximum of the perturbed species DHAP is next in amplitude followed by G3P and 3PG. G6P and F6P have within experimental error no response. In contrast to the ordering of the amplitudes, the maximum of the DHAP deviation occurs in time before that of the maximum of the F1,6BP deviation. A pulse of G3P brings a response only from DHAP. There are no responses from a pulse of 3PG.

In recording the responses from a pulse of NADH we plot absolute concentrations versus. time (fig. 6.6), since there are positive and negative responses. G3P shows a significant positive response, the remainder are negative. The ordering of extremes in time is: F1,6BP, G3P, G6P, and 3PG.

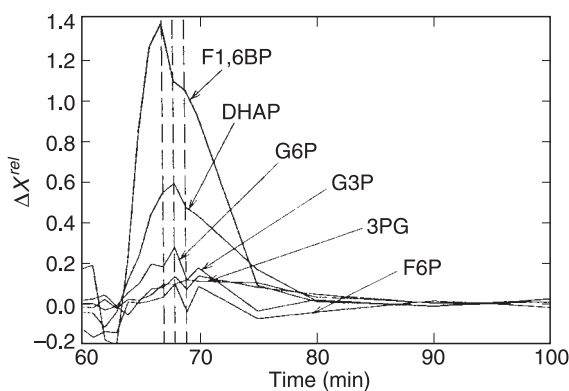


Fig. 6.4 Relative variations of several concentrations after a pulse of F1,6BP. (From [1].)

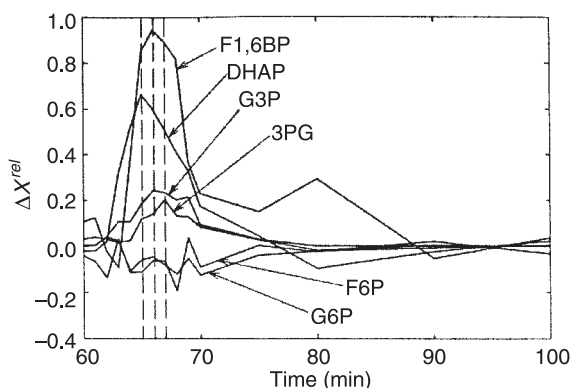


Fig. 6.5 Relative variations of several concentrations after a pulse of DHAP. (From [1].)

Now we come to an interpretation of these experimental results. The measurements are of limited precision but adequate for a useful analysis. In this analysis we made a point of sequestering ourselves as much as possible from prior available information in order to establish a severe test of the pulse method. In fact, we first assigned random numbers to the chemical species, and then repeated the analysis with the known names of the species. The conclusions were not different and hence for ease of presentation we use the chemical names. We could not detect GAP and 1,3-bisphosphoglycerate. Further, we assumed DHAP and GAP to be in rapidly established equilibrium, which we confirmed by experiment. The ratio of DHAP to GAP was found to be close to the equilibrium constant of the isomerization, 0.045. The equilibrium is way on the side of DHAP, and therefore the concentration of GAP is quite small and not measurable. A number of features of the reaction pathway can be deduced from the experiments with a pulse of G6P. Strong propagation to F6P suggests that this species follows G6P; F6P is the first species to respond to the pulse. The pulse of G6P propagates next to DHAP, which follows F6P in the pathway.

The most damped responses to a pulse of G6P were those of 3PG and G3P; these two species may therefore follow DHAP or they may participate in one or more irreversible reactions preceding G6P. We shall see that we can distinguish between these

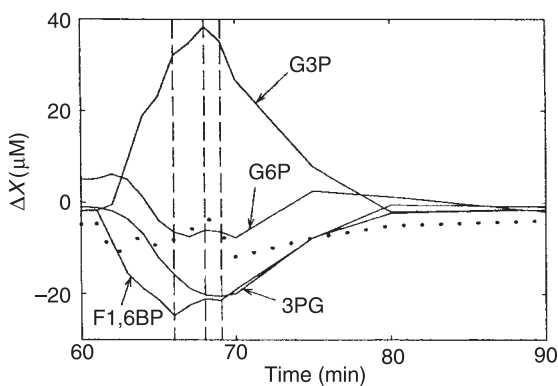


Fig. 6.6 Absolute variations of several concentrations after a pulse of NADH. The dotted line is the sum of variations of F1,6BP, G3P, and 3PG. (From [1].)



two possibilities. Pulses of F1,6BP (fig. 6.4) showed that it must be an intermediate between F6P and DHAP. DHAP displayed the strongest response to these pulses, whereas propagation to G6P and F6P is small. This result suggests that an irreversible reaction occurs between F6P and F1,6BP. Pulses of DHAP resolve further issues. The response of F1,6BP is strong and quick and the amplitude of its maximum is larger than that of DHAP. This suggests that the stoichiometry of the transition from F1,6BP to DHAP is not 1:1, that these two species are connected by fast reactions, and that the equilibrium constant, with F1,6BP as reactant, must be low (reported values are 0.069 mM). The maximum relative deviation of F1,6BP is approximately twice that of DHAP, and we conclude that two molecules of DHAP are formed from one molecule of F1,6BP, as one would conclude from a knowledge of the molecules and conservation of mass. (Compare with the example in fig. 5.7, the reaction  $2X_2 \rightarrow X_3$ , and eq. (5.17).) G3P and 3PG respond similarly, with low intensity. If these two species were located before G6P, then their responses to a DHAP pulse should not be detectable; remember that propagation of G6P pulses to DHAP is strong. The concentrations of G6P and F6P vary hardly at all, indicating again that irreversible reactions must be present. Since the response of F1,6BP to DHAP is strong, such irreversible reactions must precede F1,6BP, in agreement with the F1,6BP pulse results.

3PG does not respond to a pulse of G3P and vice versa. These species could be separated by slow and/or irreversible reactions, but this hypothesis is not compatible with experiments in which the responses of these two species are similar, for example fig. 6.5. Thus these two species have a common precursor, but neither is a precursor of the other. Hence a bifurcation of the reaction chain into two branches must occur and one species is on one branch and the other species on the other branch. There is no response of any species to a pulse of 3PG, which shows that it is a terminal species produced by either several or irreversible reactions, as these two possibilities would prevent the pulse from propagating upstream. A slight response of DHAP is obtained from a pulse of G3P, which indicates that these two species are connected by few and reversible reactions.

Additional evidence for two branches in this reaction system is obtained from experiments with pulses of NADH (fig. 6.6). The deviation in G3P is positive but that of 3PG negative. We can explain that by assuming that NADH is an activator or a substrate in the branch with G3P and an inhibitor or product in the branch with 3PG. The sum of the deviations of F1,6BP, G3P, and 3PG is close to zero (see the dotted line in fig. 6.6), which we interpret as follows: NADH elicits an overall activation on the two branches leading to 3PG and G3P which increases the sum of the concentrations of these two species. The concentration of the common precursor decreases by the same amount. DHAP follows F1,6BP and therefore DHAP is that precursor. However, the response of DHAP to a pulse of NADH was too small to measure; variations in DHAP propagate quickly to F1,6BP, and hence this species shows a negative deviation. F1,6BP and DHAP are close to equilibrium.

From all these deductions the connectivities of chemical species and a reaction pathway can be inferred (fig. 6.7). G6P is transformed into F6P, with some reversibility in that step. F6P in turn is transformed irreversibly into F1,6BP, which subsequently produces two molecules of DHAP in a fast reaction that is close to equilibrium. At this point in the reaction pathway a bifurcation occurs into two branches, one leading to G3P the other to 3PG. NADH is involved as effector or substrate on the branch producing G3P, and as an inhibitor or substrate on the branch producing 3PG. From the

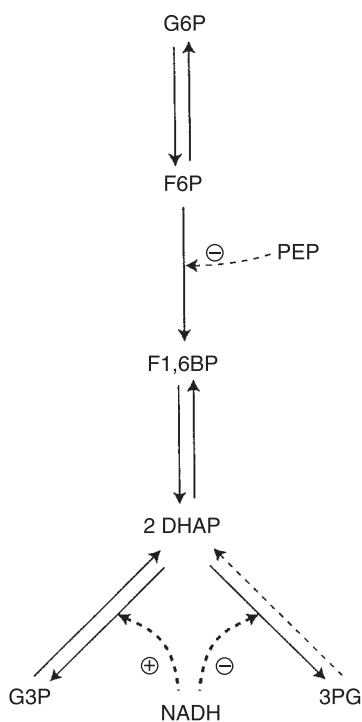


Fig. 6.7 Proposed reaction scheme based on experiments. Dashed lines with circles indicate that activation (+) or inhibition (−) maybe effected by a metabolite either as a substrate or as a product, or as an effector. PEP, phosphoenolpyruvate. (From [1].)

determined connectivities of the reaction pathway in fig. 6.7 and the known chemistry of the individual species, much of the reaction mechanism can be established. From other experiments we determined that phosphoenolpyruvate (PEP) is an inhibitor of the enzyme PFK, which catalyzes the reaction of F6P to F1,6BP.

We see that the pulse method described in the previous chapter and applied in experiments in this chapter yields causal reaction connectivities, reaction pathways, and reaction mechanisms. The global structure of the reaction network is obtained by piecing together local connectivities. The number of species to be measured depends on the system, but not necessarily all species require pulsing. The concentrations of the species need to be measured in time, but since only the location and amplitude of the maximum response of a species to a pulse is required, only a modest number of measurements are needed. Frequently reaction orders and sometimes rate coefficients also become available.

For the application of the pulse method to an unknown reaction system we would of course use all information available, which we resisted in this test example. The pulse method does not require the pulsing of each reactant species. We did not perturb F6P but showed its connectivities from pulses of G6P and F1,6BP. Further, we did not measure NADH (although it is easy to do so spectroscopically), but pulses of this species showed its connectivities in the reaction system. Reactions that are fast compared with others may be difficult to detect because responses of more than one species may occur

simultaneously. They may affect the stoichiometry, particularly at branching points, as we observed for the two molecules of DHAP produced from one molecule of F1,6BP.

*Acknowledgment* This chapter is based on the article “Experimental test of a method for determining causal connectivities of species in reactions” by Antonio Sanchez Torralba, Kristine Yu, Peidong Shen, Peter Oefner, and John Ross [1], with minor changes in wording.

## Reference

- [1] Sanchez Torralba, A.; Yu, K.; Shen, P.; Oefner, P. J.; Ross, J. Experimental test of a method for determining causal connectivities of species in reactions. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 1494–1498.

## Correlation Metric Construction

### *Theory of Statistical Construction of Reaction Mechanisms*

---

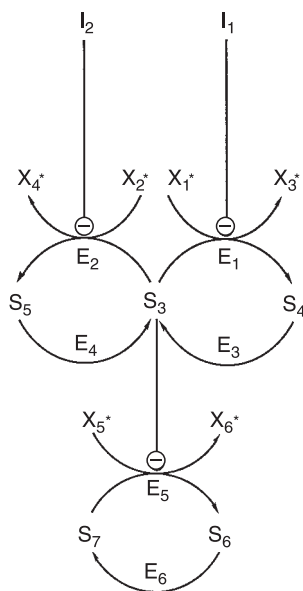
In chapter 5, we studied the responses of chemical species in a reaction system to pulse perturbations and showed the deduction of direct, causal connectivities by chemical reactions—the reaction pathway—and the reaction mechanism from such measurements. The causal connectivities give the information on how the chemical species are connected by chemical reactions. In this chapter we turn to another source of information about chemical species in a reaction system, that of correlations among the species.

Correlations of concentrations are measures of the statistical dependence of the concentration of one species on that of one or more of the other species in the system. Such correlations can be determined from measurements of time series of concentrations collected around a stationary state (nonequilibrium or equilibrium). We shall show that from concentration correlations it is possible to construct a skeletal diagram of the reaction system that gives a graphical measure of strong control and regulatory structure in reaction networks, gives some information on connectivity, leads to information on the reaction pathway and mechanism, and may simplify the analysis of such networks by identifying possible, nearly separable subsystems [1].

We begin with the demonstration and explanation of this approach by analyzing some abstract reaction models. In chapter 8 we show the utility of this “correlation metric construction” (CMC) with application to time-series measurements on a part of glycolysis, and in chapter 13 on an extensive genome study.

Consider a chemical system as shown in fig. 7.1. Mechanisms of this type are common in biochemical networks. For example, the subnetwork of fig. 7.1 containing  $S_3$  to  $S_5$  is based on a simple model of fructose interconversion in glycolysis [2] (see fig. 1.2) and the subnetwork composed of  $S_6$  to  $S_7$  is similar to the phosphorylation/dephosphorylation cycles found in cyclic cascades [3,4]. As an aside, this mechanism performs the function

Fig. 7.1 Chemical reaction mechanism representing a biochemical NAND gate. At steady state, the concentration of species  $S_6$  is low if and only if the concentrations of both species  $I_1$  and  $I_2$  are high. All species with asterisks are held constant by buffering. Thus, the system is formally open although there are two conservation constraints. The first constraint conserves the total concentration of  $S_3 + S_4 + S_5$ , and the second conserves  $S_6 + S_7$ . All enzyme-catalyzed reactions in this model are governed by simple Michaelis–Menten kinetics. Lines ending in  $\ominus$  over an enzymatic reaction step indicate that the corresponding enzyme is inhibited (noncompetitively) by the relevant chemical species. We have set the dissociation constants,  $K_{D,i}$ , of each of the enzymes  $E_1$ – $E_6$ , from their respective substrates equal to 5 concentration units. The inhibition constants,  $K_{I1}$  and  $K_{I2}$ , for the noncompetitive inhibition of  $E_1$  and  $E_2$  by  $I_1$  and  $I_2$ , respectively, are both equal to 1 unit. The  $V_{\max}$  for both  $E_1$  and  $E_2$  is set to 5 units, and that for  $E_3$  and  $E_4$  is 1 unit/s. The  $V_{\max}$ 's for  $E_5$  and  $E_6$  are 10 and 1 units/s, respectively. (From [1].)



of a biochemical NAND gate, another example of a chemical computational function (see chapter 4). The goal of CMC is to determine both the regulatory structure and the connectivity of the species in the elementary reaction steps of this mechanism, solely from measurements of the responses of the concentrations of  $S_3$  to  $S_7$  to externally imposed variations in the input concentrations,  $I_1$  and  $I_2$ . The choice of input species and response species is arbitrary but may be influenced by some prior knowledge of the system. In the system in fig. 7.1 we have a total of 7 species. We assume initially that it is possible to identify and measure all the chemical species that make up the chemical network. This is a strong assumption, but with continuing improvements in instrumentation it is increasingly satisfied. Further, we assume that we may impose concentration variations independently on a subset of chemical species in the network at each time.

If the system is initially in a stationary state, then the imposition of concentration variations of  $I_1$  and  $I_2$  will remove the system from that stationary state and, after a given variation, the system returns to the stationary state with a given relaxation time. The differences in time between adjacent concentration measurements are assumed to be on the order of, or longer than, the slowest relaxation time in the network. If the differences in time are much shorter than the relaxation time, the system cannot respond; the system will act as a low-pass filter and degrade the input signal. On the other hand, if the intervals in time are too long, then the system returns to its stationary state. Some filtering—degrading of the signal—must occur to obtain information on the correlations from the time series.

The imposed concentration variations can be chosen; for example, they can be periodic or random. We shall choose them to be random. The procedure of CMC requires

several steps which we now detail.

1. The set of measurements of the concentrations of the species in  $S_i$  is obtained as a function of the externally controlled concentrations of the species  $I_1$  and  $I_2$  at each of the selected time points. Figure 7.2 is a plot of the time series for each of the species in this system. One time point is taken every 10 s for 3,600 s. The effects of using a much smaller set of observations are discussed later. The first two plots are the time series for the two externally controlled inputs. The concentrations of  $I_1$  and  $I_2$  at each time point are chosen from a truncated Gaussian (normal) distribution centered at 30 concentration units with a standard deviation of 30 units. The distribution is truncated at zero concentration. The choice of Gaussian noise guarantees that in the long time limit the entire state-space of the two inputs is sampled and that there are no autocorrelations or cross-correlations between the input species. Thus all concentration correlations arise from the reaction mechanism. The bottom five times series are the responses of the species  $S_3$  to  $S_7$  to the concentration variations of the inputs.

It is interesting to note in fig. 7.2 that this chemical system, as in many if not most other reactions, acts a frequency filter; much of the higher frequency noise in the two inputs  $I_1$  and  $I_2$  is filtered out in the responses, the more so the further the response species is separated from the input species by reactions. For more discussion on this subject see [5].

2. From the measured (in this case calculated) time series of the concentrations in fig. 7.2 we form correlation functions. Denote by  $x_i(t)$  the concentration of species  $I_i$  at time  $t$ , and by  $\langle x_i \rangle$  the average of that concentration over the entire time of the measurements; then the time-lagged correlation function of two species  $i$  and  $j$ ,  $S_{ij}(\tau)$ , is defined by

$$S_{ij}(\tau) = \langle (x_i(t) - \bar{x}_i)(x_j(t + \tau) - \bar{x}_j) \rangle \quad (7.1)$$

for a given time lag  $\tau$ , which may be positive, negative, or zero. If the random variations of  $I_1$  and  $I_2$  affect the concentration of species  $i$  independently from that of species  $j$ , then the correlation function  $S_{ij}(\tau)$  is zero for any value of  $\tau$ . We have so far introduced only pair correlation functions for two species at a time. This measure of association is, in many ways, a precursor of later methods that utilize higher moments of the joint probability distribution among the system variables. We discuss one of these extensions at length in chapter 9 and another related approach in chapter 13 (see section 13.6, “Bayesian Networks”).

It is convenient to discuss normalized correlation functions  $r_{ij}$ , defined by the equation

$$r_{ij}(\tau) = \frac{S_{ij}(\tau)}{\sqrt{S_{ii}(\tau) S_{jj}(\tau)}} \quad (7.2)$$

The indices  $i$  and  $j$  range over all species in the system. The correlation functions depend in a complicated way on the elementary reactions and their rate coefficients.

There are two conservation constraints in this system: first, the sum of the concentrations of  $S_6$  and  $S_7$  is constant; second, the sum of the concentrations of  $S_3$ ,  $S_4$ , and  $S_5$  is constant. Hence of the 5 response concentrations, only 3 are independent; the matrix  $\mathbf{R}$  made up of the correlations  $r_{ij}$  is four-dimensional, three concentration dimensions and the time lag  $\tau$ .

Figure 7.3 is a set of plots of some three-dimensional cross-sections from the four-dimensional  $\mathbf{R}(\tau)$  surface calculated from the data in fig. 7.2. Each cross-section represents the correlations, at all calculated time lags (here every  $\pm 10$  s up to

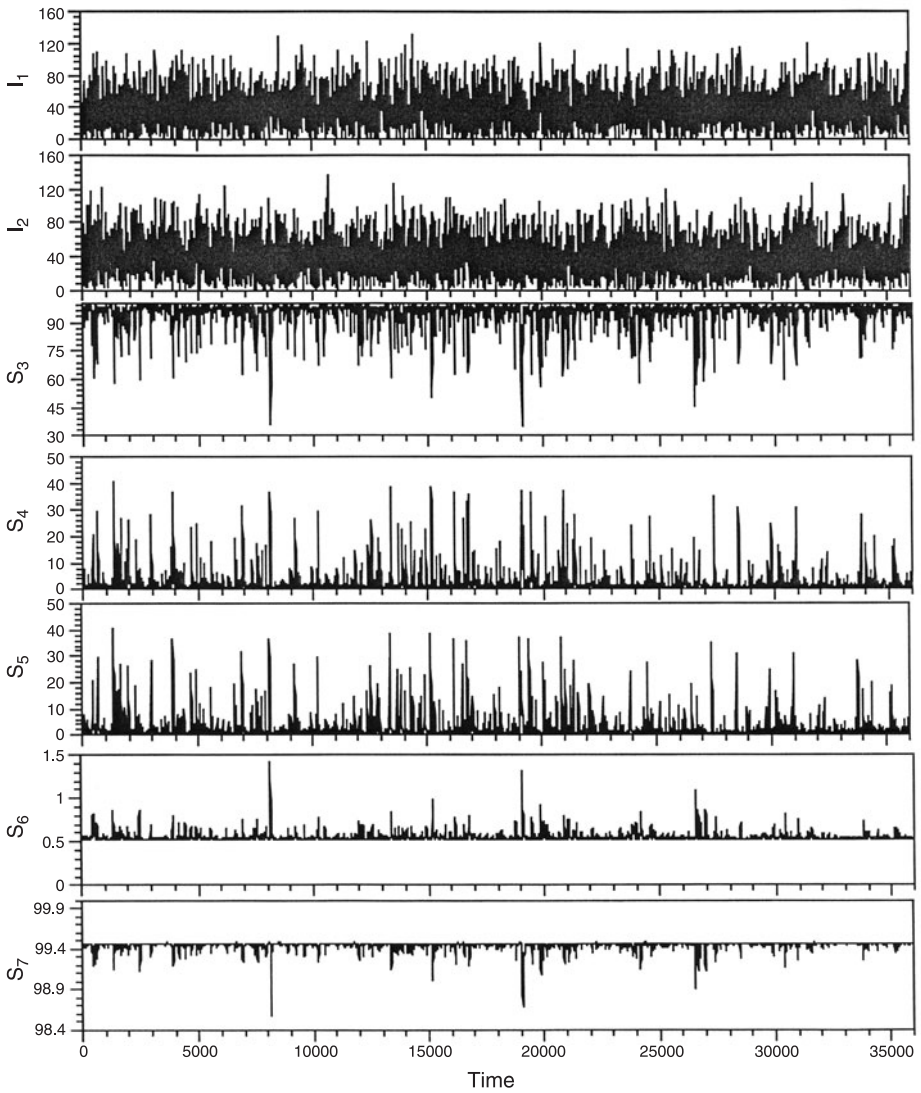


Fig. 7.2 Plot of the calculated concentration time series for all the species composing the mechanism in fig. 7.1. Only the first two time courses (those for  $I_1$  and  $I_2$ ) are set by the experimenter. The concentrations of  $I_1$  and  $I_2$  are chosen independently from a Gaussian distribution with a mean and standard deviation of 30.0 concentration units. Since the lower limit of concentration is zero, the actual distribution of input concentrations has a tail toward high concentrations. See step 1 in the text for a full explanation. (From [1].)

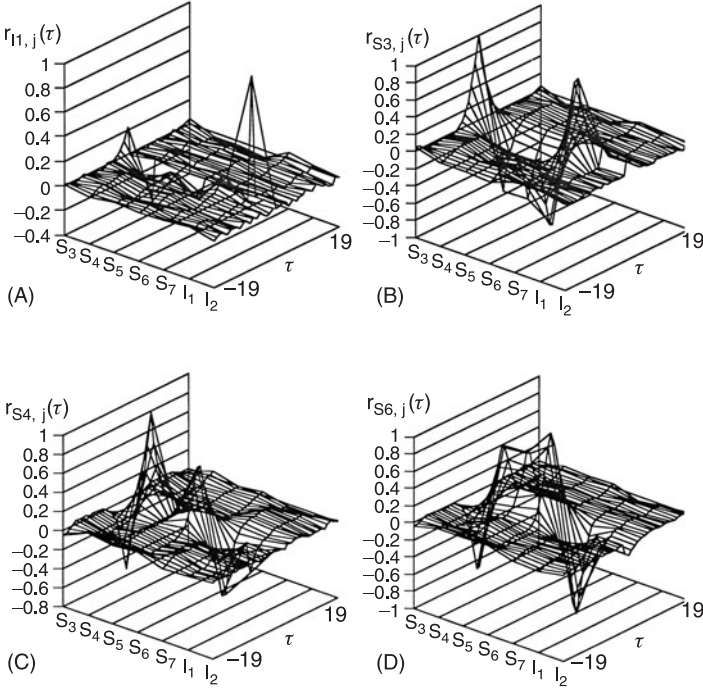


Fig. 7.3 Plots of cross-sections through the four-dimensional time-lag correlation surface calculated from eqs. (7.1) and (7.2). The cross-sections are A =  $r_{I_1, j}(t)$ , B =  $r_{S_3, j}(t)$ , C =  $r_{S_4, j}(t)$ , and D =  $r_{S_6, j}(t)$ . See step 2 in the text for explanation. (From [1].)

a time lag of  $\pm 190$  s), of the times series corresponding to a given species with those of each of the other species, and itself. If the system were truly at steady state at each time point, then the correlation surface for combinational networks (those in which there are no feedback loops) would be flat except on the  $\tau = 0$  plane. Figure 7.3 shows only four cross-sections corresponding to the choice of three independent species and one of the inputs. Examination of the cross-sections of  $\mathbf{R}(\tau)$  (fig. 7.3D) shows that, as expected, the concentration of  $S_6$  perfectly anticorrelates with that of  $S_7$  (by conservation) and negatively correlates with that of  $S_3$ .

Figure 7.4 shows two projections of the cross-sections in fig. 7.3B. Figure 7.4A is a projection on the species correlation plane. This projection shows the relative correlations of each of the species with  $S_3$ . Figure 7.4B is a projection on the time-lag correlation plane. Here, a rough sequence of events in time becomes clear when it is noted that if one species correlates with a second species at a positive lag, then the variation of the first species occurs (on average) before that of the second. The correlation of a species with itself at nonzero lag is expected due to relaxation to the steady state and should be symmetric around zero lag. The sequence of events with respect to  $S_3$  is therefore as follows:  $I_1$  and  $I_2$  are changed externally, followed by the simultaneous changes in  $S_3$  to  $S_5$ , after which  $S_6$  and  $S_7$  change. In this case, this time line is a sequence of causal connectivity. However, in branched networks or networks with feedback, causal connectivity may not always be determined in



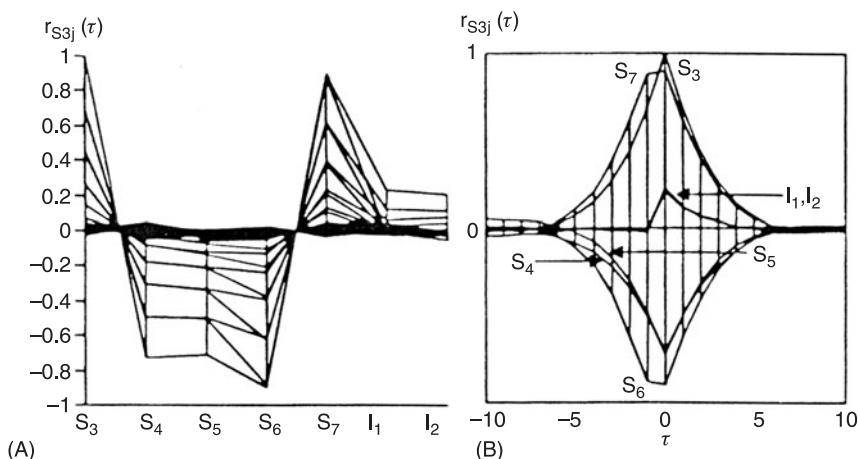


Fig. 7.4 Projections of fig. 7.3B down the species (A) and time-lag (B) axes. Figure 7.4A emphasizes the correlation of  $S_3$  with all the other species. It correlates most highly with  $S_6$  and  $S_7$ , next with  $S_4$  and  $S_5$ , and finally with  $I_1$  and  $I_2$ . Figure 7.4B (truncated to emphasize the region between  $\tau = \pm 10$  units) shows the large amount of time information that is largely ignored by the simple analysis presented here. If a species correlates at a positive lag with respect to  $S_3$ , then that species leads  $S_3$  in time; a negative lag implies that the correlating species follows  $S_3$ . Since fig. 7.4B shows that the correlation with the inputs tails toward positive lags, action on the inputs precedes action at  $S_3$ . Action at  $S_4$  and  $S_5$  occurs simultaneously with  $S_3$ , and  $S_6$  and  $S_7$  follow action at  $S_3$  (as indicated by their left tails). In this particular network this sequence of lags relates directly to the causal sequence. The inputs affect the concentration of  $S_3$ – $S_5$ , which then affects the concentration of  $S_6$  and  $S_7$ . The amount of this information available for a given experiment is dependent on the magnitude of the rate coefficients with respect to the time between measurements. (From [1].)

this way. Further, if the time between measurements had been 70 s rather than 10 s, the network would have been at steady state and no time delays would be observed.

Causality, in this case the assignment of which reactants lead to which products, cannot be determined simply by the strength of a statistical relation. Three criteria must be met for causality to be established: there must be (1) a temporal ordering of the variables; (2) concomitant variation; and (3) control over other factors that might affect the observed relationships among the variables. In CMC we do not always meet the first criterion, and the third criterion can only be fulfilled by data outside the present analysis, for example, knowledge of the basic chemistry. This point is especially relevant to the connection algorithm discussed in step 3.

The correlation matrix is symmetrical with 1's on the diagonal. If two or more covariates are perfectly correlated, this implies that there will be columns and rows of the correlation matrix that are exactly the same. Thus the correlation matrix will not be of full rank. In noiseless experiments, perfect correlation can arise for a number of reasons. The two most common would be the existence of a conservation condition among two or more species and the establishment of quasi-equilibria. In the case of conservation, the decrease in one concentration is necessarily followed by the increase in others. For two species in a conservation relationship this results in perfect anticorrelation between the variates. For quasi- or actual equilibria the

input signal has effectively propagated instantaneously throughout the system; thus, since the final concentration of each species in the “equilibrium relationship” will be a function only of the kinetic parameters and the current value of the inputs (but not time), their dynamics will be perfectly correlated (unless there is no variation in one of the concentrations). In our examples there are a number of such conditions.

The nullity (the number of zero eigenvalues) and null-space (the space spanned by the corresponding eigenvectors) of the correlation matrix determine the number of conservation constraints and rapidly established quasi-equilibria in the network and the species involved, respectively (since species involved in such relations are completely dependent on one another). For example, the eigenvalues of the (zero lag) correlation matrix derived from the time series in fig. 7.3 are  $A = (3.89, 1.3, 0.97, 0.67, 0.17, 1.7 \times 10^{-7}, 7.0 \times 10^{-9})$ . Thus, since the last two eigenvalues are so small compared to the first five, the nullity of the matrix is 2, as expected from the two conservation conditions. The correlation matrix can be cleaned up somewhat by zeroing all correlations that are not significant by the standard  $t$ -statistic-based significant tests [6].

3. We suggest a connection algorithm which generates an approximate dependency list among the different species. Ideally, we would like the dependency list to reflect directly the reactions among the species. However, dependency as obtained from the correlation studies is not necessarily related to causal connectivity, but rather to a high degree of association among sets of variables. There are a number of common analyses, such as multiple regression, canonical correlation [7], linear structural relation (LISREL or SEPATH) [8,9], and Box–Jenkins time-series analysis [10], which attempt to determine the dependencies of a set of observable dependent variables on known independent variables. All of these techniques are essentially regression analyses in which an estimate of the mean of one set of variables is made based on knowledge of another set of variables.

In each of the cases above, a general model structure is asserted. An example is a simple nonlinear regression (as discussed in chapter 1) in which the observation of one variation at time  $t$  is dependent on other variables at time  $t - \tau$ . A simple regression model might look like eq. (1.1), where the regression methods solve for the  $\beta$ 's which are conceptually (and sometimes actually) related to our correlation measures of association. This example is relatively unstructured and methods like LISREL allow the assertion of more causal hypotheses into the regression. One goal of CMC is to construct an approximate model of the interactions among species, without explicit specification of a model (chemical mechanism). Correlation-based analysis measures a symmetric degree of association between two variables. Thus, correlations yield a convenient metric of distance between two species. Since the balance of our method (steps 4–7) relies on such a metric, we develop a simple correlation-based agglomerative dependency algorithm with the following iterated steps that are very similar to the procedures one would use to hierarchically cluster the data:

- (a) Place each species in its own group.
- (b) Find the two groups,  $i$  and  $j$ , each containing a disjoint subset of species for which the magnitude of correlation (at any lag) between a species in  $i$  and one in  $j$  is the maximum over all pairs of groups. That is, find the maximum correlation between two species *not* in the same group.
- (c) Merge these two groups into one group and list the connection, found in step (b), between the two species (one from each original group) with the maximum correlation. If two or more species from one group correlate with species in the other with (nearly) the same maximum magnitude, then list connections

between these as well. All listed connections are designated as “significant” connections, since we did not reject the correlation by the  $t$ -statistic mentioned above.

- (d) Go back to (b), now with one fewer group, and repeat the steps until there is only a single group left. This procedure creates a singly linked system graph in which every species is connected to at least one other species.

Application of the connection algorithm to the full correlation surface of the NAND mechanism yields the values of the significant connections listed in table 7.1. The algorithm neglects the direct effect of each input species on the concentration of  $S_3$  since we are using only a single-link dependency; the magnitude of the correlation between the inputs and  $S_3$  is weak since  $S_3$  is maximally affected only when the inputs are both low or both high at the same time (a relatively rare event). It may be possible to eliminate such statistical misses and redundancies in the connection list by including chemical knowledge, incorporating the available correlation time-lag information, and application of the transinformation criterion of probabilistic reconstruction analysis, a variant of which we use in EMC and ERT, discussed in chapter 9, which uses conditional probabilities derived from the full (calculated) joint probability density function generated from the time series. Lastly, the difference in correlations of  $S_4$  and  $S_5$  to  $S_3$  is merely a result of the sampling statistics. There is also a standard  $t$ -statistic to test for the significance of the difference for two correlations drawn from the same experimental data.

The connections among species, since they are based on correlations, represent a noncausal structural model of the system. Correlations may be decomposed into four parts: (1) direct effects, that is, causal connectivity, in which one variable is a direct antecedent to a second (e.g., one species is directly converted to another by a chemical reaction); (2) indirect effects, where one variable influences another by way of a third (e.g., one species affects the production of a second which is then converted directly to a third); (3) spurious effects, which occur when two variables have a common antecedent (e.g., when one species is converted into two other species by separate reactions); and (4) unanalyzed effects, which arise from correlations between the externally controlled variables [11]. In order to construct a mechanism (a causal model), only the first contribution to the correlation between two variables should be considered. We do not here separate the measured correlation into these explicit components. In the absence of special knowledge of the chemical mechanism, methods for doing so, such as path analysis and LISREL [8,9], require extensive and complex calculation as well as a number of restrictive assumptions. Our simple definition of connection, though possibly not as informative as a full analysis of causal connectivity, yields a good first guess for a relational structure among the species (ideally defined by the first component of correlation above).

Table 7.1 Listing of the “significant” connections calculated among the chemical species composing the reaction mechanism in fig. 7.1

	$S_4$	$S_5$	$S_6$	$S_7$
$I_1$	-0.31			
$I_2$		-0.32		
$S_3$	-0.72	-0.71	-0.90	0.90
$S_6$				-1.00

A partial description of the functions of the network may be read from the signs of the connections in table 7.1. For example, since  $I_1$  and  $I_2$  are negatively correlated with  $S_4$  and  $S_5$ , both of which are negatively correlated with  $S_3$ , we may hypothesize that  $S_3$  is high only when  $I_1$  or  $I_2$  or both are high, and  $S_3$  is low otherwise. So the subsystem of fig. 7.1 composed of  $S_3$ – $S_5$ , and with the output defined as  $S_3$ , may be functionally analogous to a logical OR (or an AND) gate between  $I_1$  and  $I_2$ . In this particular case, the function is an AND gate. If only  $I_1$  is high, then material partitions between  $S_3$  and  $S_5$ , and if only  $I_2$  is high, then material partitions between  $S_3$  and  $S_4$ . If  $I_1$  and  $I_2$  are both high, then all materials funnel toward  $S_3$ . If the kinetics are set up correctly (reaction from  $S_3 \rightarrow S_4, S_5$  fast without  $I_{1,2}$  and very slow with  $I_{1,2}$ ), then this arrangement accumulates  $S_3$  only when  $I_1$  and  $I_2$  are both high.

4. The time-lagged correlation matrix,  $\mathbf{R}(\tau)$ , is converted into a Euclidean distance matrix with the canonical transform [7]:

$$d_{ij} = (c_{ii} - 2c_{ij} + c_{jj})^{1/2} = 2^{1/2} (1.0 - c_{ij})^{1/2} \quad (7.3)$$

$$c_{ij} = \max |r_{ij}(\tau)|_\tau \quad (7.4)$$

where the second equality in eq. (7.3) follows from the properties of the correlation matrix. The formula in eq. (7.4) defines  $c_{ij}$  to be the absolute value of the maximum correlation between the time series for species  $i$  and that of species  $j$ , regardless of the value of  $\tau$ . The neglect of the value of  $\tau$  at the maximum correlation between two time series represents a loss of useful information. In kinetic systems in which the correlation is peaked at some nonzero lag, the implication is that the system is large enough, in the sense of sufficient species, for there to be a delay time between the site of signal generation (the inputs) and the reception. This information may be incorporated into the distance metric in order to represent better the distances among the species. In addition, the sign of the lag may be used to estimate a sequence of events.

We define  $\mathbf{D} = (d_{ij})$  to be the distance matrix. Since  $\mathbf{D}$  is Euclidean, its elements automatically satisfy the three standard tests for a metric space: identity, symmetry, and the triangle inequality. In the case of perfect correlation between two variables, the triangle inequality is violated, a situation that may be remedied (without dire consequences) by adding a small value,  $\varepsilon = 1 \times 10^{-10}$ , to the distance between them. The particular metric defined by eq. (7.3) is a measure of independence between two variables. If the correlation between two variables is small, then the distance between them is large.

5. The classical multidimensional scaling (MDS) method [7,12] is applied to the distance matrix calculated in step 4 in order to find both the dimensionality,  $\Delta$ , of the system and a consistent configuration of points representing each of the species. Before explaining the mathematics of that process, let us give a pictorial representation. Think of each of the significant connections in table 7.1 as represented by a stick: the length of the stick is determined by the absolute magnitude of the correlation listed in the table and the stick carries a label of one species at one end and that of another species at the other end. For example, for the correlation of species  $S_3$  and  $S_5$  the stick has a length of 0.71 and is marked at one end with  $S_3$  and at the other end with  $S_5$ . Now take the 7 sticks corresponding to the 7 entries in the table, pick out all those with  $I_1$  on one end, and place those ends at one point. Next, do the same for all ends marked  $I_2$ , and so on. You will need a multidimensional space to accomplish this, but you will have constructed a multidimensional object. Shine a beam of light on the object and observe the projection of the object on a screen.

Rotate the object until the information that you have about the object on the screen is a maximum. For the case of the object built from the entries of correlations in table 7.1 the best projections are given in fig. 7.5, which displays well the reaction pathway of the mechanism in fig. 7.1.

To accomplish the task just described mathematically, we use a standard approach by defining a centered inner product matrix  $\mathbf{B}$  by the equations

$$\mathbf{B} = -(1/2)\mathbf{H}(d_{ij})\mathbf{H} \quad (7.5)$$

$$\mathbf{H} = \mathbf{I} - (1/M)\mathbf{I}\mathbf{I}' \quad (7.6)$$

where  $\mathbf{H}$  is the centering matrix [7],  $\mathbf{I}$  is the  $M \times M$  identity matrix, and  $\mathbf{I}'$  is the  $M \times M$  unit matrix. The operation of the symmetric, idempotent matrix  $\mathbf{H}$  on the vector  $\mathbf{x}$  has the effect of subtracting off the mean of the entries of  $\mathbf{x}$  from each of its elements, that is,  $\mathbf{H}\mathbf{x} = \mathbf{x} - \bar{x} \mathbf{I}$ , where  $\bar{x} = n^{-1} \sum x_i$ . The number of significant eigenvalues of  $\mathbf{B}$  (defined in table 7.2) is the dimensionality,  $\Delta$ , of the system, and the vectors constructed from the first  $\Delta$  coordinates of the  $M$  eigenvectors comprise the principal coordinates of points representing the chemical species in the correlation diagram. (Formally, each point represents the particular time series generated for a given chemical species.) The eigenvectors are normalized such that  $\mathbf{z}_i \cdot \mathbf{z}_i = \lambda_i$ . The distance between each pair of points is inversely related to the correlation between the corresponding species. If the  $M$  series are independent, then all the  $d_{ij}$  are equal to  $2^{1/2}$  and points representing each species must fall on the vertices of a regular  $M - 1$ -dimensional hypertetrahedron. At the other extreme, when all species perfectly correlate (or anticorrelate), then the  $d_{ij}$  are all equal to 0 and there is a single degenerate point in the system. By construction, however, the inputs are completely uncorrelated, so the minimum dimension derived from a CMC is  $|I| - 1$ . Most often, we are interested in the first two principal coordinates of the MDS solution since the configurations may then be plotted on a plane and are thus easy to visualize.

The numerical results from the MDS analysis of the distance matrix derived from eqs. (7.3) and (7.4) are shown in table 7.2. The columns of table 7.2A are the eigenvectors of  $\mathbf{B}$ , and the rows are the coordinates of each point (time series). Each eigenvector,  $\mathbf{z}_k$ , corresponds to the projection of the  $k$ th coordinate of each of the points on an orthogonal basis vector. Each of the eigenvalues (table 7.2B) is an indicator of the degree to which the vectors from the origin of the configuration to each point are projected along the corresponding basis vector. In this example, over 99% (for  $a_{2,3}$ ) of the distance matrix is "explained" by the first three eigenvalues. Thus, we find the dimensionality  $\Delta \approx 3$ , which represents a reduction of three dimensions from the theoretical maximum of six. Two of these dimensions are lost due to the two conservation constraints. The third dimension is lost due to constraints imposed by the degree of interaction among the subsystems (the conservation conditions lead to a degeneracy in the correlation matrix). The two-dimensional projection of the coordinates of the points listed in table 7.2A is shown in fig. 7.5A. Each pair of points with a nonzero entry in table 7.1 is connected by a line. This represents the correlation metric diagram which depends on the reaction mechanism and rate coefficients (as well as the properties of the perturbations, such as the frequency spectra, average values, and standard deviations). This diagram reproduces many features of the standard mechanistic diagram (fig. 7.1) but

Table 7.2 Eigenvectors and eigenvalues of the matrix **B** (eq. 7.5) determined by the classical MDS analysis of a distance matrix calculated from the correlation surface in fig. 7.3 (see steps 4 and 5 in the text)

(A) Eigenvectors,  $z_k$ :

Point/ $z_k$	1	2	3	4	5	6	7
1 ( $I_1$ )	6.68e-01	-5.84e-01	4.05e-01	5.51e-02	1.77e-02	-8.20e-09	-3.52e-10
2 ( $I_2$ )	7.00e-01	5.26e-01	-4.30e-01	4.93e-02	1.42e-02	-8.20e-09	-3.52e-10
3 ( $S_7$ )	-4.20e-01	7.29e-03	-8.16e-03	2.05e-01	1.90e-03	-6.82e-09	-7.67e-09
4 ( $S_6$ )	-4.20e-01	7.29e-03	-8.16e-03	2.05e-01	1.90e-03	-9.58e-09	6.97e-09
5 ( $S_4$ )	-1.44e-01	-5.51e-01	-4.02e-01	-1.60e-01	-7.55e-02	-8.20e-09	-3.52e-10
6 ( $S_5$ )	-7.15e-02	5.60e-01	4.30e-01	-1.38e-01	-7.65e-02	-8.20e-09	-3.52e-10
7 ( $S_3$ )	-3.14e-01	3.49e-02	1.27e-02	-2.16e-01	1.16e-01	-8.20e-09	-3.52e-10

(B) Eigenvalues,  $\lambda_k$ :

$k$	$\lambda_k$	$a_{1,k}$	$a_{2,k}$
1	1.413496e+00	3.978311e-01	4.937104e-01
2	1.237497e+00	7.461268e-01	8.721279e-01
3	6.958617e-01	9.419783e-01	9.917824e-01
4	1.805556e-01	9.927960e-01	9.998381e-01
5	2.559576e-02	1.000000e+00	1.000000e+00
6	4.747935e-16	1.000000e+00	1.000000e+00
7	1.080736e-16	1.000000e+00	1.000000e+00

The coefficients  $a_{1,k}$  and  $a_{2,k}$  are agreement measures of the degree to which the distance matrix is “explained” by the  $k$ -dimensional MDS solution. The measures are calculated from

$$a_{1,k} = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n |\lambda_i|} \quad \text{and} \quad a_{2k} = \frac{\sum_{i=1}^k \lambda_i^2}{\sum_{i=1}^n \lambda_i^2}$$

where the eigenvalues are sorted in decreasing order. Thus, according to  $a_{2,k}$ , 49.4% of the distance matrix is explained by one dimension ( $a_{2,1} = 0.494$ ) and 100% of the matrix is explained by a configuration in five dimensions ( $a_{2,5} = 1.0$ ). Columns 1 and 2 of the eigenvector matrix define the  $x$  and  $y$  positions of points in fig. 7.4A, respectively.

differs in that it has additional information, such as the extent of coupling (tightly or loosely) among subsystems in the network.

6. Figure 7.5A is a projection of the high-dimensional MDS object onto two dimensions. The distances between the points in the 2D representation are therefore less than or equal to actual measured distances. A more representative 2D diagram may often be obtained with an optimization-based MDS method that allows the distances between the optimized points to be greater than, equal to, and less than the actual distances (see [11]). The diagram derived with this algorithm from our example is shown in fig. 7.5B, which is more representative of the actual distance

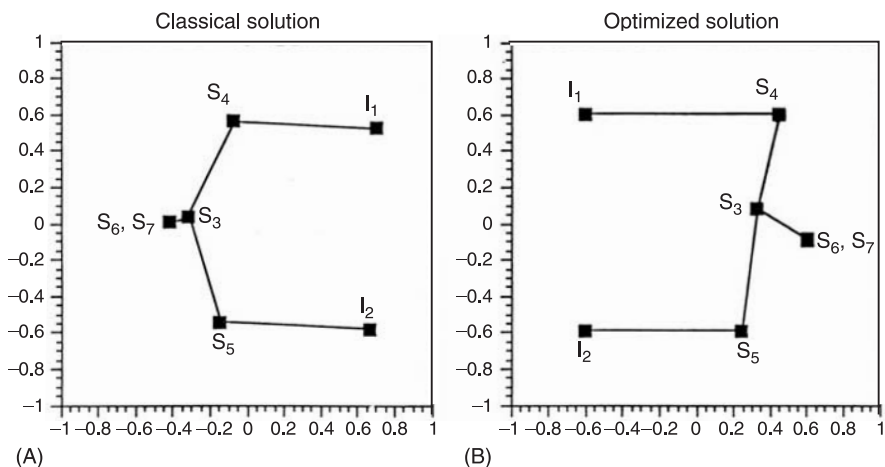


Fig. 7.5 Classical and optimized MDS solutions calculated from the canonically transformed correlation surface from fig. 7.3. (A) The two-dimensional projection of the three-dimensional ( $a_{2,3} = 99.2\%$ ) (or four-dimensional;  $a_{2,4} = 99.98\%$ ) object found by the matrix method described by step 5 in the text. The coordinates for the points in A are the same as the first two columns of the matrix of eigenvectors in table 7.2A. The second diagram (B) is derived from the MDS optimization method discussed in step 6. The lines between points in both diagrams are obtained from the results of step 3 (table 7.1). Thus, B is more representative of the measured distance matrix. Both diagrams correspond to rotated and slightly distorted approximations to the reaction mechanism in fig. 7.1. (From [1].)

matrix (see the caption of fig. 7.5). Note that all rotations, reflections, and translations of the MDS diagrams are equally valid MDS solutions. Both diagrams, parts A and B of fig. 7.5, are rotated and slightly distorted versions of the diagram of the reaction mechanism shown in fig. 7.1. Thus, from the state measurements (calculations) of the time series, we derive a construction related to the reaction mechanism for the system, but with the added information on the relative coupling strengths among species.

7. Finally, a cluster analysis is performed on the distance matrix. This method is used to summarize the grouping of chemical subsystems within the reaction mechanism and to give a hierarchy of interactions among the subsystems. There are many techniques of cluster analysis [7,12,13], but for simplicity we employ a nonparametric hierarchical clustering technique called the weighted pair group method using arithmetic averages [14]. The algorithm operates on  $\mathbf{D}$  in three iterated steps:

- Search the  $M \times M$  distance matrix for the minimum distance between pairs of clusters and let this distance be  $d_{ij}$ . In the initial step, there are  $M$  clusters ( $u_1, u_2, \dots, u_M$ ) each containing a single point (species).
- Define a new cluster,  $u_k$ , containing the two objects ( $i$  and  $j$ ) with the minimum distance between them. Define the branching depth of this cluster as  $h_{ij} = d_{ij}/2$ . Finally, define the distance between the new cluster,  $k$ , and all other clusters,  $l \neq i$  or  $j$ , by the equation  $d_{kl} = (d_{il} + d_{jk})/2$ .

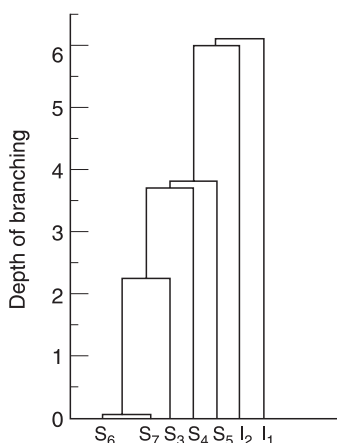


Fig. 7.6 Hierarchically clustered dendrogram calculated from the CMC analysis of the mechanism in fig. 7.1. The dendrogram is produced as described in step 7 in the text. The hierarchy is a good representation of the flow of control from the inputs down to  $S_6$  and  $S_7$ . (From [1].)

- (c) Delete clusters  $i$  and  $j$  from the distance matrix and add the newly computed cluster distances to the matrix. If  $M$  is not equal to 1, then decrease  $M$  by one and return to step 1, otherwise the algorithm is done.

This procedure generates a dendrogram. Such dendrograms have also been used to visualize and aid interpretation of the results of microarray experiments [15]. The dendrogram derived from a cluster analysis of the distance matrix calculated in step 4 is shown in fig. 7.6. The tree may be interpreted as follows. Since the branching depths of species  $I_1$  and  $I_2$  are nearly the same, but they do not belong to the same cluster, we may consider the two as separate subsystems which couple with similar strengths to the rest of the network (the species below the inputs in the diagram). A similar assumption may be made about species  $S_4$  and  $S_5$ . The diagram is then simple to analyze. Since we control species  $I_1$  and  $I_2$ , these in turn most strongly influence the changes in concentrations of both  $S_4$  and  $S_5$ . These two species then combine to control the concentrations in the subnetwork ( $S_3$ ,  $S_6$ ,  $S_7$ ). This three-member subnetwork may be further divided into  $S_3$  and ( $S_6$ ,  $S_7$ ). From the correlation metric diagram and the eigenanalysis of the correlation matrix we know that  $S_4$  and  $S_5$  are the determinants of  $S_3$  (since the connections are significant and the three species are involved in a conservation relation), which subsequently controls the concentrations of  $S_6$  and  $S_7$ . These last two species also satisfy a conservation constraint. Thus, a hierarchical diagram of control is derived.

We next discuss some more examples in order to clarify the usage and interpretations of CMC analyses. Figure 7.7 shows a chemical system composed of two subsystems, one like that of fig. 7.1 (subsystem 1) and another realizing a NOT  $I_1$  and NOT  $I_2$  function (subsystem 2). The kinetics of subsystem 1 are chosen to be faster than the analogous system in fig. 7.1. Subsystem 2 is composed of substrate cycles similar to those in subsystem 1, but the kinetics of the enzyme reactions are much slower than that of subsystem 1. We thus introduce two different time scales for the two subsystems.



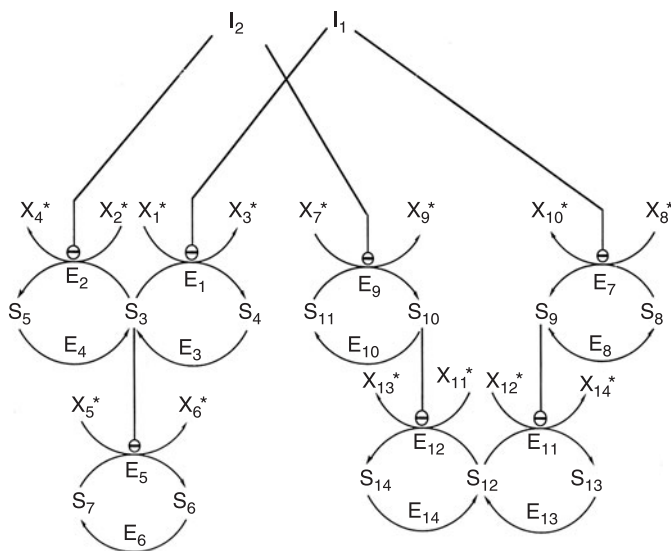
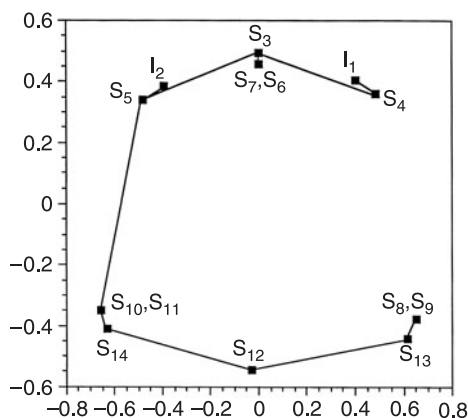


Fig. 7.7 Mechanism composed of two different realizations of a chemical NAND gate. The first NAND-like submechanism is composed of species  $S_3$ – $S_7$  plus the inputs. The second is composed of  $S_8$ – $S_{14}$  plus the inputs. Both subsystems have the inputs as common causal antecedents. As in fig. 7.1, the concentrations of all species bearing an asterisk are considered to be held constant by buffering or external flows. (From [1].)

This system was chosen to demonstrate the concept of chemical subsystems as defined by CMC and to demonstrate two interesting interrelated problems that arise during the analysis: (1) ambiguity arising from common causal antecedents and (2) low-pass filtering of the input signal(s). Figures 7.8 and 7.9 show the results of the CMC analysis of this network. The two subsystems separate onto the upper and lower half-planes of fig. 7.8 and onto two different branches of the dendrogram in fig. 7.9, respectively. The MDS configurations of the two subsystems have similar structures, as expected. The input species  $I_1$  and  $I_2$ , however, group very close to  $S_4$  and  $S_5$  on the far side of the MDS diagram away from  $S_{13}$  and  $S_{14}$ . This occurs

Fig. 7.8 Two-dimensional projection of the classical MDS solution resulting from a CMC analysis of the mechanism in fig. 7.7. Note that, despite the fact that both NAND gate subsystems of the mechanism are driven by common inputs, the different subsystems group on different half-planes of the diagram. The placement of the inputs of the diagram is due to the choice of rate constants (see text). The reasons for the connection drawn between ( $S_{13}$ ,  $S_{14}$ ) and  $S_5$  and for the lack of connection between  $S_{11}$ ,  $S_{12}$  and the other subsystem or inputs are described in the text. For this diagram,  $a_{2,4} > 99.8\%$ . (From [1].)



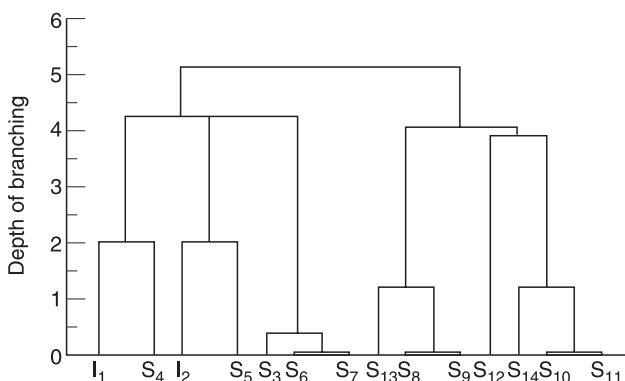


Fig. 7.9 Hierarchically clustered dendrogram calculated from the CMC analysis of the mechanism in fig. 7.7. The two different NAND subsystems separate onto the two major branches of the dendrogram. The inputs cluster with the first subsystem (the one similar to the mechanism in fig. 7.1) since its kinetics are rapid compared to the second subsystem. The branch containing subsystem 1 is not structurally equivalent to the dendrogram in fig. 7.6 because, as a result of the kinetics of subsystem 1 being fast compared to those of the mechanism in fig. 7.1, the inputs couple more tightly to  $S_4$  and  $S_5$ . (From [1].)

because the rates of interconversion of, for example,  $S_3$  and  $S_4$  are faster than the corresponding interconversion of  $S_8$  and  $S_9$ ; the concentration of species  $S_4$  is able to follow better the fluctuations in the input species than  $S_8/S_9$ . The  $S_3/S_4$  dynamics is more responsive to  $I_1$  than  $S_8/S_9$ , which is slow enough to significantly filter the input perturbations. So  $I_1$  seems more correlated with  $S_3/S_4$  than  $S_8/S_9$ .

If the enzymatic reactions are slow, then not very much material is converted each time the inputs change state. Large changes in concentration are only obtained (assuming Gaussian driving noise) when there are substantial low-frequency components in the input signals. Thus, slow reactions act as low-pass filters for their input signals. Though both the subnetworks  $S_3$ – $S_5$  and  $S_8$ – $S_{11}$  filter the signals sent by changes in concentration of  $I_1$  and  $I_2$ , examination of the Fourier transforms (not shown) of the time series for each set of species shows that the slower kinetics of  $S_8$ – $S_{11}$  leads to a much stronger exponential decay of the components in their frequency spectra than those of  $S_3$ – $S_5$ . The correlation (which is related to the product of the Fourier transforms of two time series) of  $I_1$  with  $S_8$  is therefore much less than with  $S_4$ . The fact that both  $S_3$ – $S_5$  and  $S_8$ – $S_{11}$  filter the signals (albeit to different extents) implies that subsystem 2 is better correlated with  $S_3$ – $S_5$  than with  $I_1$  and  $I_2$ . This also explains why  $I_1$  and  $I_2$  appear on the far side of subsystem 1 with respect to subsystem 2. It may be possible to correct for the filtering effects by weighting the Fourier transforms of the series by an appropriate exponential factor. This may allow better decomposition of the correlation into its direct, indirect, and spurious components. It must be remembered, however, that CMC relies on the fact that each chemical mechanism filters input signals in a characteristic fashion, resulting in an identifiable MDS diagram. This point is emphasized again in the next example.

The formulation of many reaction mechanisms is frequently based on one of two simplifying possibilities [16]: (1) The reaction mechanism has one rate-determining step. (2) There is no rate-determining step, but the concentrations of intermediates

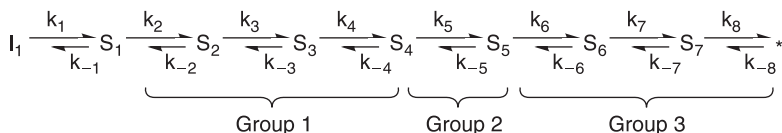


Fig. 7.10 Linear reaction network. All reactions are first-order mass action kinetics. The back reactions ( $k_{-1}$  to  $k_{-8}$ ) are always  $0.1 \text{ s}^{-1}$ ;  $k_1$  is set to  $1.0 \text{ s}^{-1}$ . The remainder of the forward rate constants are broken into three groups, as shown. Within a group, all the forward coefficients are assumed to be identical. These rate constants are chosen from three possible rates: 0.7, 70.0, and  $7,000.0 \text{ s}^{-1}$  (slow (S), medium (M), and fast (F)). (From [1].)

are (nearly) constant (stationary state hypothesis). Figure 7.10 shows a simple unbranched chain of chemical reaction steps. Such conversions are common structures in metabolic pathways, and thus provide an interesting case study for CMC. In order to demonstrate the effects of different patterns of rate coefficients on the outcome of a CMC analysis, we break the linear network into three sets of reactions: those governed by the first three rate coefficients,  $k_2$ – $k_4$ , are the first group;  $k_5$  defines the second group; and the last three rate coefficients define the third. Each group of coefficients may be defined to be slow ( $0.7 \text{ s}^{-1}$ ), medium ( $70 \text{ s}^{-1}$ ), or fast ( $7,000 \text{ s}^{-1}$ ) with respect to the switching time of the input time series in  $I_1$  (switching frequency =  $20 \text{ s}^{-1}$ ). Figure 7.11 shows MDS diagrams resulting from CMC analysis of the linear networks with five different sets of rate coefficients. In all cases, time series were collected every  $0.05 \text{ s}$  for  $100 \text{ s}$ . The average concentration of the input was one unit, and the standard deviation was  $0.1$  units.

Figure 7.11(a) shows the results from a scheme in which all three groups of rate constants are chosen to be slow (no rate-determining step). When all the rate coefficients in a set of consecutive reactions are comparable, the stationary state hypothesis is often employed to solve analytically the kinetic equations. The approximate autocorrelation time for the time series, defined as the time lag at which the correlation of a series with itself decays to zero, is about  $2.5 \text{ s}$  (50 lag times, i.e., 50 observations). The system is not very close to steady state in this case, but a reasonably linear MDS diagram is produced nonetheless. There are three points to be made:

1. The large distance between the input and  $S_1$  occurs because fluctuations in the input species occur too quickly for  $S_1$  to follow. Thus, the filtering effect described above leads to a larger distance than expected.
2. The distance between adjacent species decreases with the number of steps away from the input.
3. The diagram is a curve instead of a line.

These last two points are related. The distance between adjacent species decreases because of filtering. As the signal propagates through the network, the high-frequency components get successively filtered out until only the frequencies slower than the characteristic relaxation times of the network survive. As filtering becomes more severe in the network, subsequent species become more strongly correlated (since they are able to follow one another exactly). This same effect is partly responsible for the curvature of the line. Since later species become more correlated, their correlation with earlier species also becomes more similar. That is,  $S_6$  and  $S_7$  are highly correlated, which in this case implies that they are correlated at nearly the same (relatively low) level with  $S_1$ . Therefore, the point representing  $S_7$  must be

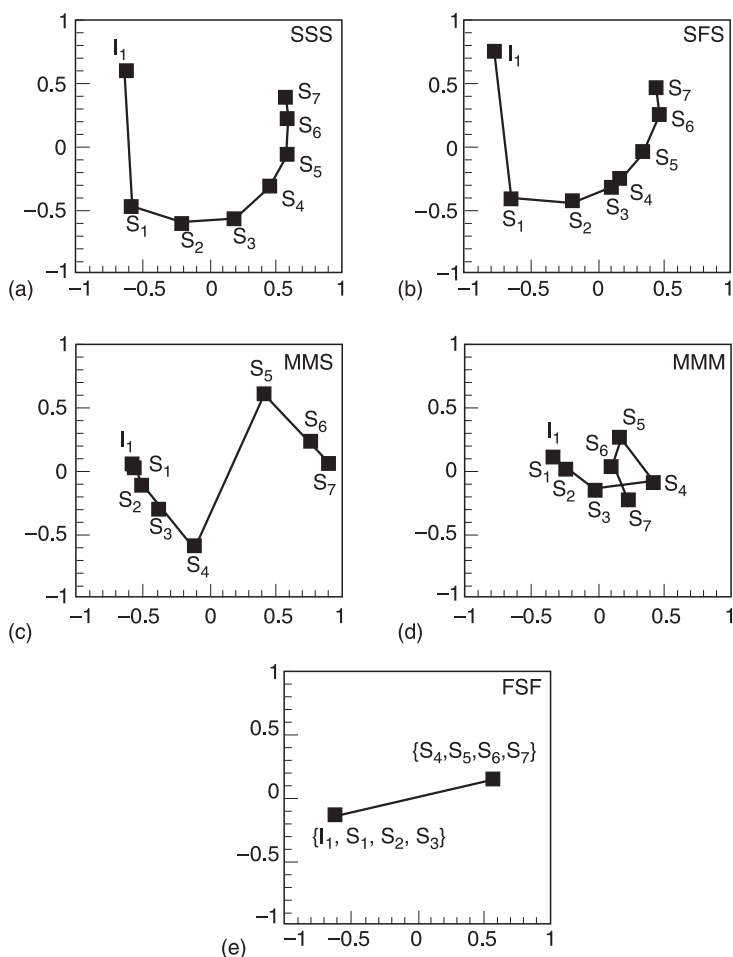


Fig. 7.11 Two-dimensional projection of the MDS solutions of the five kinetic schemes for the linear mechanism in fig. 7.9. The three-letter triad in the upper right corner of each diagram indicates the pattern of rate constant for the first, second, and third parts of the network, respectively. See the caption of fig. 7.10 and the text for more information. (From [1].)

placed at approximately the same distance from  $S_1$  as  $S_6$ . Further, since lower correlations are measured less accurately than high correlations (see below), the difference between  $r(I_1, S_6)$  and  $r(I_1, S_7)$  becomes less significant (not as accurately measured). This leads to the so-called horseshoe effect often seen in the MDS of serial data [7].

Figure 7.11(b) shows the case in which the middle step is fast compared to the other reaction steps. Since the rate of conversion of  $S_4$  to  $S_5$  is large, we expect the distance between these two species to be small. However, in the diagram it is the distance between  $S_3$  and  $S_4$  that is markedly decreased. Since  $S_4$  is immediately converted to  $S_5$ , which itself is slowly degraded, the autocorrelation time of  $S_5$  increases. This implies that its correlation with  $S_4$  (the autocorrelation of which

relaxes immediately) is weakened. However, since the back reactions are small, the (relatively) small concentration of  $S_4$  follows  $S_3$  nearly perfectly.

Figure 7.11(c) shows the case in which the reactions in the last part of the network are slow. Here, the distance between  $S_4$  and  $S_5$  is large, as expected. Since this distance is largely a result of a severe low-pass filtering of the signal sent by  $S_4$ , the correlation of  $S_4$  with all species in the last part of the network is relatively low. Thus, these last species are almost equidistant to  $S_4$  even though the distances among them are significant. This results in the observed displacement of the  $S_5$  from  $S_7$  away from the rest of the network.

Figure 7.11(d) is the most pathological of the diagrams. It shows a loop in the structure of the network. The rate constants here are all medium. The autocorrelation time in this network is on the order of the fluctuation time of the input (0.05 s). With these rates, a signal initiated at the inputs travels approximately halfway down the network by the time of the next input concentration. This implies, for example, that at zero time lag  $S_1$  correlates highly with the input and species  $S_2$  and  $S_3$  but not with  $S_5$ – $S_7$ .  $S_1$  correlates with these latter species at one lag time. Since  $S_4$  is observed just as it is responding to the current signal but still relaxing from its response to the last signal, its correlation with  $S_1$  at both lags is smaller than expected. Since the correlation metric does not take into account the time at which the maximum correlation of each species with  $S_1$  occurs,  $S_5$  appears “closer” to  $S_1$  than does  $S_4$ . This leads to the observed loop in the diagram. If the metric is modified to take into account only the zero lag correlation, then the expected linear diagram is obtained (data not shown). This example shows how the characteristics of the input noise can strongly affect the outcome of a CMC experiment in some regimes. However, such pathologies can often be diagnosed by direct examination of the correlation surface. For example, maximum correlation at significant time lags implies intermediate steps in the mechanism and the decay of correlations in time can detect fast reactions.

Finally, fig. 7.11(e) shows the case in which all reactions are fast except the one in the second group which is slow. This corresponds to the case of a network with a rate-limiting step. Here, the autocorrelation times in the network are much faster than the input fluctuations, so all species in the first and last section of the network are effectively at steady state and therefore perfectly correlated. The slow step in the middle, however, causes filtering of the signal and thereby causes a decrease in correlation between the two sections of the network. This results in the dumbbell shape of the network.

## Discussion

With the application and development of concepts of electronic circuit and systems theory, and techniques from multivariate statistics, we obtain a MDS correlation metric diagram that defines and represents the putative (not necessarily causal) connections given by the correlations, and the strength of kinetic interactions among the species of a reaction network from a time-series determination of chemical concentrations. In addition, we obtain the startling result that, for the systems studied, the MDS diagram recapitulates many features of the reaction mechanism.

A number of caveats, improvements, and new directions are apparent. Correlation metric construction is a statistical analysis, hence care must be exercised to take into account the significance of all the calculated correlations. If we ignore any error in the

concentration measurements themselves, the probability that  $|r_{ij}(\tau)|$  is larger than it should be in the null hypothesis (which is that species  $i$  and  $j$  are uncorrelated) is  $\text{erfc}\{[r_{ij}^2(\tau)n/2]^{1/2}\}$ , which is a rapidly decreasing function of  $n$ , the number of observations. Smaller correlations are therefore less significant for a given  $n$  than large correlations, a fact that is currently ignored by our connection algorithm. This may lead to the so-called horseshoe effect often seen when MDS is used to seriate data [7]. Experiments might be continued until none of the correlations are changing by more than a given percentage. However, we must be aware that for a given set of parameters, defining the random perturbations on the inputs, and for a finite number of measurements the calculated correlations are only valid for the portion of the state-space of the system observed. The state-space of a reaction network is defined here by the range of all possible sequences of input concentrations parametrized by the initial values of the concentrations of all the other species composing the network.

Once the correlations are properly measured, the rest of the analysis proceeds simply, except for the calculation of the connection diagram. The connections among points in the MDS solution determine which species participate in each reaction step or interaction (for example, one species might be an effector for the enzymatic production of another species). The connection algorithm described in step 3 is primitive in that (1) causality, that is, causal connectivity, cannot be hypothesized since even the value of  $\tau$  at the maximum correlation between two species is ignored (and see discussion above), and (2) the minimum correlation sufficient for connection between two species is chosen merely on the assumption that all species are connected to at least one other species. This algorithm implies a somewhat arbitrary definition of significance of correlation between a single variable and one or more other variables. It may be possible that transinformation [17], partial correlation analysis, and multiple regression or path analysis [7–9,17,18] can be used to discover high-order relations not initially uncovered by the relatively simple correlation analysis.

One such example of a high-order relation is the combined effect of  $I_1$  and  $I_2$  on  $S_3$  in fig. 7.1. Transinformation, multiple regression, and LISREL also provide measures of significance via the  $\chi^2$  statistic and provide estimates of the others. These methods may therefore lead to a better definition of the connection among species as well as providing an indication of the existence of unmeasured variables in the case when the variation in a measured species cannot be explained by variations in itself, the other measured species, and the inputs. Suppose, for example, we can measure three species in a network,  $I_1$ , A, and B. Suppose further that  $I_1$  is a direct precursor of both A and B, but that it must react with an unknown (or not directly measurable) species U in order to produce B. Assuming the concentration of U varies significantly during the time course of the experiment, part of the fluctuation of (B) is due to these changes in (U). It is, then, impossible to explain fully the variance in B using the two measured variables  $I_1$  and A and prior values of B, and as a result, a significant residual,  $e_B$ , will be found when B is regressed on these known variables. A large residual between the measured variance of B and the amount of variance of B explained by the known variables thus indicates the existence of unmeasured species.

The MDS correlation diagram is a new representation of a reaction network. Aside from reproducing in large part the mechanistic diagram for a chemical reaction network, information is present regarding tightness of kinetic coupling among the species. For example, species  $S_6$  and  $S_7$  in the mechanism in fig. 7.1 are drawn as far apart as species

$S_3$  and  $S_4$ . However, in the diagrams plotted in fig. 7.5,  $S_6$  and  $S_7$  are plotted on top of one another since they are completely correlated. Thus, the MDS diagram is distorted to reveal loci of tight regulation; for example, on the MDS diagram the point representing  $S_3$  is close to the pair  $S_6$  and  $S_7$  and is the species whose concentration most determines their concentration. The hierarchical clustering results, then, are largely a recapitulation of what is evident by eye from the MDS diagram.

There is a related point: in the example derived from fig. 7.1, the concentrations of the input species are within the region of saturation of their target enzymes (concentrations much greater than  $K_1$ ) most of the time; however, these concentrations may take on values near or below the inhibition constant, and thus transitions between low and high concentrations of  $S_3$  are observed. If we had chosen the parameters of the Gaussian driving noise such that the input concentrations were always much greater than the inhibition constants, then such a transition would never be observed and the correlation between  $S_3$  and the pair  $S_6$  and  $S_7$  would have been much reduced. Alternatively, if the average values of the input concentrations had been in the transition region of enzymes (close to  $K_I$ ), the response of the species to variation of the inputs would be highly nonlinear. In this case, linear correlation coefficients would underestimate the relationship between variables. In general, Spearman nonparametric rank correlation [6,19] should be used, since by using the rank order of correlations rather than their absolute values it accurately measures not only linear relations among variates but also monotonic ones.

Besides the average concentration and standard deviation, the other important parameter of the driving noise on the inputs is the frequency spectrum of the external perturbations. If the concentrations of the inputs switch at a rate comparable to the slowest relaxation times in the system, then the network will attenuate the high-frequency components in the input sequence. In our analysis we chose the time between changes of input concentrations such that the system was driven near its steady state. Figure 7.4B shows that the species have completely relaxed in less than 10 time lags (100 s) after perturbation. This implies that if the variation of the input concentrations is much faster than 100 s (say on the order of 1 s), the network will not have time to respond and the correlation strength between system components and the inputs will be lost. This effect is a mixed blessing for the current analysis since it leads to an artificially high correlation between species at loci remote from the entry points of the inputs; the input signal may be filtered similarly at disparate but equally remote sites. The resultant correlations tend to group highly filtered species together far from those species more directly connected to the inputs. This is a desirable feature except that the distances between the highly filtered species will be smaller than expected even when such species do not interact. Further, as shown in the analysis of the mechanism in fig. 7.7, filtering may result in the spurious assignment of connection between two species.

Clearly, the final MDS diagram is partially dependent on the parameters of the noise imposed on the system. It is possible that frequency domain approaches to time series analysis [10] may help in a study of the role of frequency transfer functions in the control of chemical networks. We have assumed that all species involved in the mechanism may be identified and measured. For systems with many species this may be difficult. When there are missing species, CMC may still be performed on the measurable subset of species. The effects of the other species are subsumed into the correlations among the known species, and a consistent diagram can be constructed. The MDS diagram, then, may not be an obvious representation of the underlying mechanism. In fact, due

to signal loss in the network, certain connections among known species may be lost. On the other hand, unknown species that participate in the dynamics of the network, but are not ultimately determined by the input concentrations (i.e., unknown species produced and degraded by uncontrolled sources in the network), can be helpful in strengthening the correlations among (chemical material) to be processed by the network.

Finally, the networks analyzed in this chapter are combinational networks, that is, networks with no (explicit) feedback loops and, therefore, no memory or autonomous dynamics. Nonzero correlations away from  $\tau = 0$  are, therefore, caused only by slow relaxation of the chemical species to their steady states (slow reaction steps). In sequential systems, in which feedback exists, nonzero time-lagged correlations may be indicative of species involved in a feedback relation. For systems that contain feedback in such a way as to generate multistability and oscillations, it may be impossible, in the absence of any prior knowledge, to predict in advance how many states are available to the network and how they are triggered. However, a series of experiments has been suggested for such systems from which the essentials of the core mechanism containing feedback may be deduced (see chapter 11). The methods discussed here may be useful complementary approaches to determining reaction mechanisms of coupled kinetic systems.

*Acknowledgment* This chapter is based on selections from the article “Statistical construction of chemical reaction mechanisms from measured time series” by Adam Arkin and John Ross [1], with additional explanations and changes in wording.

## References

- [1] Arkin, A.; Ross, J. Statistical construction of chemical reaction mechanisms from measured time series. *J. Phys. Chem.* **1995**, *99*, 970–979.
- [2] Arkin, A.; Ross, J. Computational functions in biochemical reaction networks. *Biophys. J.* **1994**, *67*, 560–578.
- [3] Schacter, E.; Chock, P. B.; Stadtman, E. R. Regulation through phosphorylation/dephosphorylation in cascade systems. *J. Biol. Chem.* **1984**, *259*, 12252–12259.
- [4] Schacter, E.; Chock, P. B.; Stadtman, E. R. Energy consumption in cyclic phosphorylation/dephosphorylation cascade. *J. Biol. Chem.* **1984**, *259*, 12260–12264.
- [5] Samoilov, M.; Arkin, A.; Ross, J. Signal processing by simple chemical system. *J. Phys. Chem. A* **2002**, *106*, 10205–10221.
- [6] Ott, L. *An Introduction to Statistical Methods and Data Analysis*; Duxbury Press: Boston, 1984.
- [7] Mardia, K. V.; Kent, J. T.; Bibby, J. M. *Multivariate Analysis*; Academic Press: San Francisco, 1979.
- [8] Li, C. C. *Path Analysis: A Primer*; Boxwood Press: Pacific Grove, CA, 1986.
- [9] Hayduk, L. A. *Structural Equation Modeling with LISREL: Essentials and Advances*; John Hopkins University Press: Baltimore, 1987.
- [10] Box, G. E. P.; Jenkins, G. M. *Times Series Analysis: Forecasting and Control*; Holden-Day: San Francisco, 1976.
- [11] Dillon, W. R.; Goldstein, M. *Multivariate Analysis: Methods and Applications*; Wiley: New York, 1984.
- [12] Shepard, R. N. Multidimensional scaling, tree-fitting, and clustering. *Science* **1980**, *210*, 390–398.
- [13] Corana, A.; Marchesi, M.; Martini, C.; Ridella, S. Minimizing multimodal functions with the simulated annealing algorithm. *ACM Trans. Math. Software* **1987**, *13*, 262–281.
- [14] Swofford, D. L.; Olsen, G. J. In *Molecular Systematics*; Hillis, D. D., Moritz, C., Eds.; Sinauer Associates: Sunderland, MA, 1990; p 411.



- [15] Eisen, M. B.; Brown, P. O. DNA arrays for analysis of gene expression. *Methods Enzymol.* **1999**, *303*, 179–205.
- [16] Berry, R. S.; Rice, S. A.; Ross, J. *Physical Chemistry*, 2nd ed.; Oxford University Press: Oxford, 2000.
- [17] Broekstra, G. C-analysis of C-structures: representation and evaluation of reconstruction hypotheses by information measures. *Int. J. Gen. Sys.* **1981**, *7*, 33–61.
- [18] Marriot, F. H. C. *The Interpretation of Multiple Observations*; Academic Press: San Francisco, 1979.
- [19] Otnes, R. K.; Enochson, L. *Applied Time Series Analysis: Basic Technique*; Wiley: New York, 1978.

## Experimental Test and Applications of Correlation Metric Construction

---

In this chapter we present an experimental test case of the deduction of a reaction pathway and mechanism by means of correlation metric construction from time-series measurements of the concentrations of chemical species [1]. We choose as the system an enzymatic reaction network, the initial steps of glycolysis (fig. 8.1). Glycolysis is central in intermediary metabolism and has a high degree of regulation [2]. The reaction pathway has been well studied and thus it is a good test for the theory. Further, the reaction mechanism of this part of glycolysis has been modeled extensively [3].

The quantity and precision of the measurements reported here are sufficient to determine the matrix of correlation functions and, from this, a reaction pathway that is qualitatively consistent with the reaction mechanism established previously. The existence of unmeasured species did not compromise the analysis. The quantity and precision of the data were not excessive, and thus we expect the method to be generally applicable.

This CMC experiment was carried out in a continuous-flow stirred-tank reactor (CSTR) (see fig. 6.2). The reaction network considered consists of eight enzymes, which catalyze the conversion of glucose into dihydroxyacetone phosphate and glyceraldehyde phosphate. The enzymes were confined to the reactor by an ultrafiltration membrane at the top of the reactor [4]. The membrane was permeable to all low molecular weight species.

The inputs are (1) a reaction buffer, which provides starting material for the reaction network to process, maintains pH and pMg, and contains any other species that act as constant constraints on the system dynamics, and (2) a set of “control species” (at least one), whose input concentrations are changed randomly every sampling period over the course of the experiment. The sampling period is chosen such that the system almost,

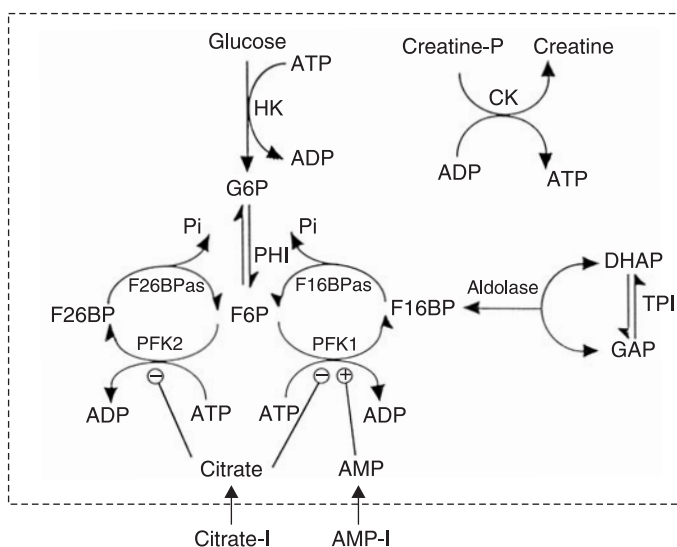


Fig. 8.1 The first few reaction steps of glycolysis. Lines ending in circles denote regulatory interaction; minus signs indicate a negative effector; plus signs, a positive effector. Creatine-P and CK keep the concentrations of ATP and ADP constant. Abbreviations for the enzymes: HK, hexokinase; PHI, phosphoglucose isomerase; PFK1, phosphofructokinase 1; F16BPas, fructose 1,6-bisphosphatase; PFK2, phosphofructokinase 2; F26BPas, fructose 2,6-bisphosphatase; TPI, triose-phosphate isomerase; CK, creatine kinase. Abbreviations for the chemical species in the system: AMP-I, input adenosine monophosphate; citrate-I, input citrate; G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; F16BP, fructose 1,6-bisphosphate; F26BP, fructose 2,6-bisphosphate; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde phosphate;  $P_i$ , inorganic phosphate; creatine-P, creatine phosphate. The remaining symbols are defined in the caption to fig. 6.1. (From [1].)

but not quite, relaxes to a chosen nonequilibrium steady state. The system is kept near enough to its steady state to minimize trending (caused by the relaxation) in the time series, but far enough from the steady state that the time-lagged autocorrelation functions for each species decay to zero over three to five sampling periods. This long decay is necessary if temporal ordering in the network is to be analyzed.

We used capillary zone electrophoresis (CZE) [5] to measure accurately the concentrations of most of the small molecular species that are passed through the membrane. Because the enzymes are not present, the concentrations in aliquots taken from the reactor outflow just before the end of each sampling period reflect the concentrations inside the reactor. The set of perturbing input species concentrations and system responses form an ordered time series suitable for analysis by CMC.

We measured the concentrations of  $P_i$ , G6P, F6P, F16BP, F26BP, and DHAP, as well as the input and reactor concentrations of citrate and AMP. We developed a CZE detection method specifically for the separation and quantification of the metabolites in glycolysis [6]. Using 5.5 mM 4-hydroxybenzoate (pH 11.8) as background electrolyte and indirect ultraviolet detection, we analyzed simultaneously the components mentioned above and

several others in an enzyme buffer. The analysis of a single sample took less than 6 min to complete. The detection limit for each species was defined as a signal-to-noise ratio of 3:1 and was determined to be 0.003 mM for  $P_i$ , 0.001 mM for F16BP and F26BP, and 0.002 mM for G6P, F6P, DHAP, citrate, and AMP. The error in repeated measurements of standard concentrations of these species was 2% for  $P_i$ , 1% for F16BP, 7% for F26BP, 2% for G6P, 3% for F6P, 1% for DHAP, 2% for citrate, and 3% for AMP.

To begin a CMC experiment, the CSTR was primed by flowing in the glycolytic enzymes to the following total activities: 4.0 units of HK, 10 units of PHI, 0.96 units of PFK1, 0.25 units of F16BPase, 0.01 units of PFK2::F26BPase, 0.86 units of aldolase, 10 units of TPI, and 75 units of CK. These activities were chosen to yield physiologically relevant concentrations of the intermediates, with the exception of F26BP, the concentration of which was set high enough to be measurable by our CZE method.

The input flow into the reactor (held constant at 0.2 ml/min) contained enzyme buffer, which includes, among other things, 0.6 mM glucose and 2.5 mM creatine-P. Glucose is the feed-reactant from which the process of glycolysis is initiated. The combination of CK and creatine-P was included to fix the ATP:ADP ratio during the course of the experiment. The reactor steady state reached when the inflow contains only these species is defined as the reference state; the intermediate concentrations were 1.48 mM  $P_i$ , 0.0086 mM F26BP, 0.20 mM F16BP, 0.072 mM F6P, 0.31 mM G6P, and 0.23 mM DHAP.

As input species, we choose two well-known indicators of cell nutrition level, AMP and citrate. AMP is an activator of PFK1 and promotes glycolysis [2], whereas citrate is an inhibitor of both PFK1 and PFK2 and inhibits glycolysis [2]. Two 55-point, uniformly random, uncorrelated time series of concentration were generated by computer: one series for AMP over the range 0 to 0.1 mM, and one for citrate over the range 0 to 1.2 mM. These ranges represent the extreme “physiological” concentrations attained by these species.

At the start of the experiment, the system was placed in the reference state. While conserving the volume of the inflow we then flowed the first set of input concentrations into the reactor. For each set of inputs in the time series, the system was allowed to relax for 10 min, and an aliquot (0.1 ml) was collected from the outflow. Hydrolysis of the sugars, including F26BP, was minimal as long as the samples were either immediately analyzed or kept frozen. The system was allowed to relax for an additional 3 min before the input concentrations were changed and the process repeated (fig. 8.2). This 13 min sampling period was chosen to be much longer than the residence time of the reactor (about 6 min) but, as determined empirically, not so long for the system to relax completely to steady state after an average perturbation. For the analysis of these measurements we calculated correlations functions as described in chapter 7. Other methods of kinetic analysis have also used a correlation measure of relatedness among reacting species [7].

A representative three-dimensional (3D) section of the entire correlation functions calculated from eq. (7.1) is shown in fig. 8.3A. This section shows the lagged correlations of the time series of all species in the system with G6P. The plot shows strong positive and negative dependencies of G6P variation on a number of other species and the inputs. The smooth decline, with increasing absolute time lag, of the autocorrelation of G6P from its maximum at  $\tau = 0$  implies that the system is relaxing to its stationary state, but is not completely at that state. Thus, some temporal ordering of variation in each of the variables may be assigned (fig. 8.3B). If the time series for a given species has

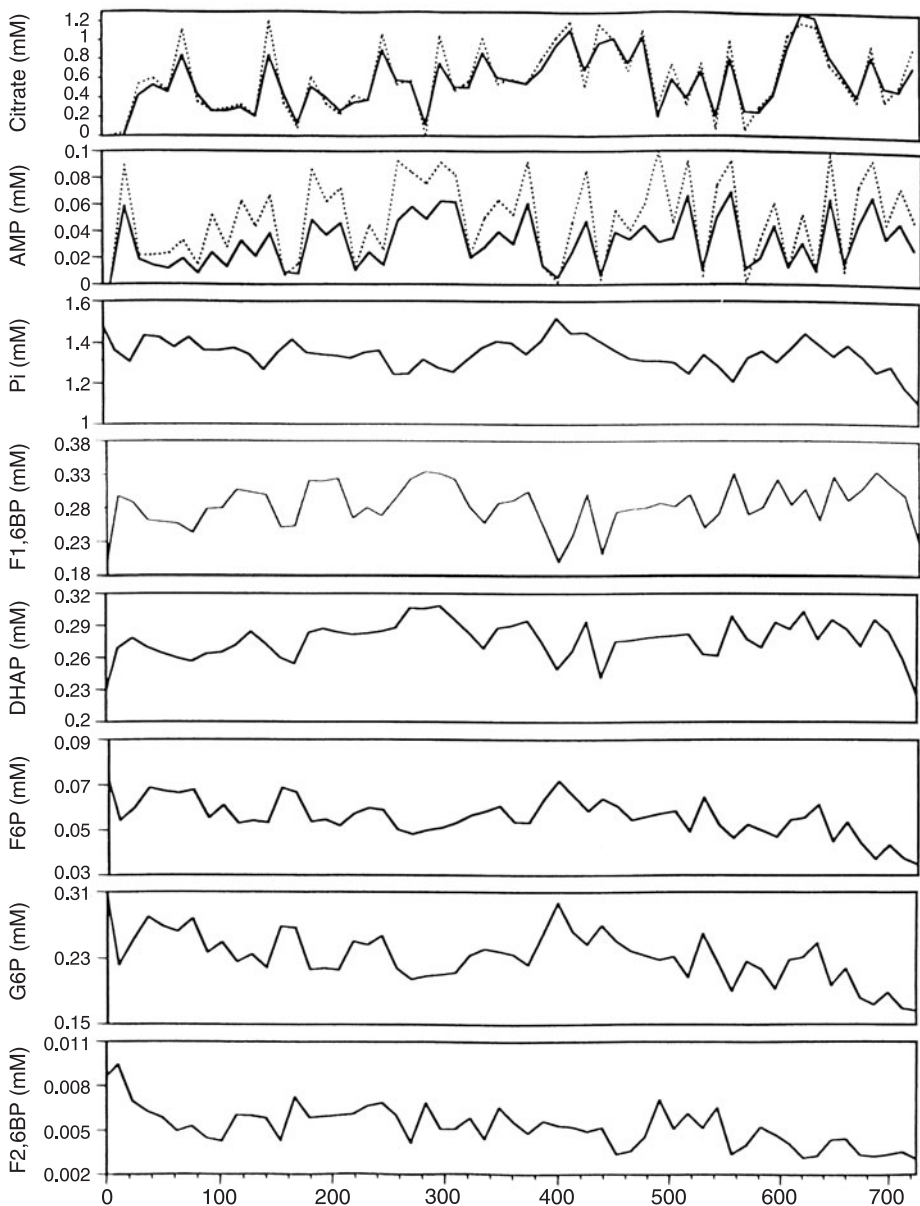
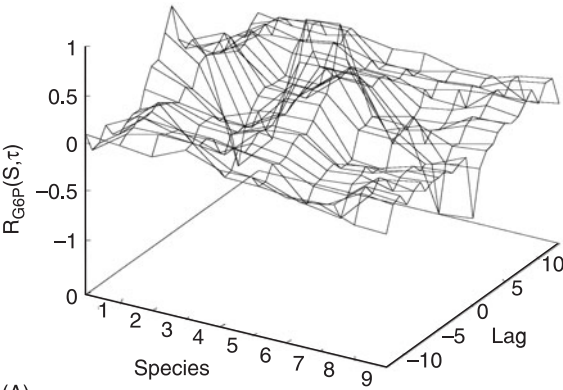
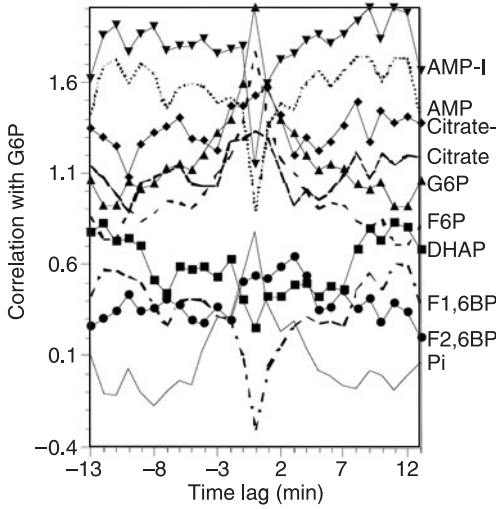


Fig. 8.2 The time courses of measured concentration of the inputs, AMP and citrate, in the experiments, with the responses of the concentrations of  $P_i$  and of the species F16BP, DHAP, F6P, G6P, and F26BP. The dotted curves are the known input concentrations for AMP and citrate. (From [1].)



(A)



(B)

Fig. 8.3 (A) Plot of the time-lagged correlation function  $R$  of G6P with all other species. Each lag represents a time of 13 min. (B) Graphs of the lagged correlation functions in (A) offset to distinguish individual correlation functions. Each successive correlation function, starting with F26BP, is offset by 0.2 unit on the correlation axis. The graph clarifies the temporal ordering data inherent in the correlation functions. For example, the maximum correlation with F26BP occurs at significantly positive lags, implying that variation in F26BP precedes variation in G6P. (From [1].)

a maximum correlation at negative lags compared to a reference time series, then that species receives the input signals after the reference series, and vice versa. Similarly, if the two series are maximally correlated at zero lag but correlation tails to negative (positive) lags, variation in the given species closely follows (precedes) variation in the reference species. For example, the tailing to positive lags of the correlation of G6P with AMP-I implies that variations in AMP-I precede variation in G6P, that is, variations in G6P are responses to variations in AMP.

A graphical matrix summary of the distance, connection, and temporal data derived from the correlation functions is shown in fig. 8.4. A network hypothesis for the system is derived simply by drawing arrows, in the directions indicated by the plus and minus symbols, between all connected species. For example, using the third row in the figure one would draw a line from F16BP to DHAP, since F16BP precedes (minus sign) DHAP, and from AMP to F16BP, since F16BP follows (plus sign) AMP. The matrix also gives

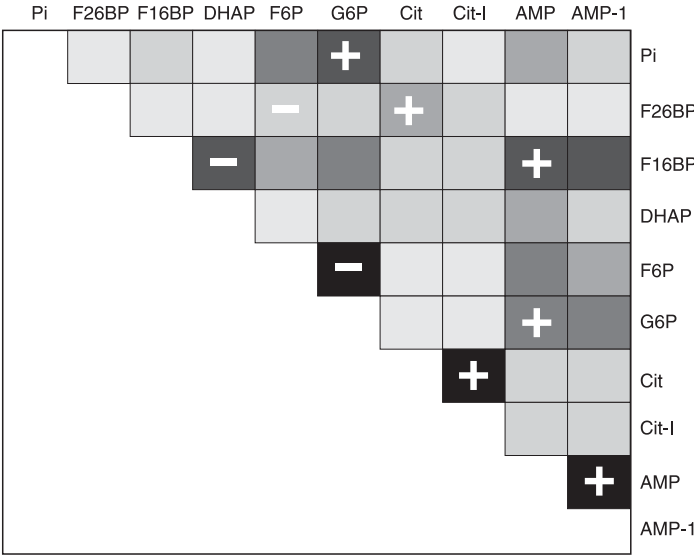


Fig. 8.4 Matrix summary of distance matrix, connection algorithm, and temporal ordering algorithm. The shade of each matrix element represents the distance between two species, calculated from the correlation function matrix by means of eqs. (7.3) and (7.4). The darker the shade, the smaller the distance (white,  $d_{ij} = 1.3$ , black,  $d_{ij} = 0.2$ ; linear gray scale). A plus or minus sign within a matrix element denotes that the connection algorithm has assigned a significant connection between these two species: further, a plus (minus) sign indicates that response of the row species follows (precedes) variation in the column species. (From [1].)

a rough idea of how “central” each species is to the dynamics of the network. For example, as indicated by the number of dark matrix elements in their rows and columns, we conclude that AMP is a much stronger effector of network dynamics than is citrate.

The annotated MDS (multidimensional scaling) diagram is automatically constructed from the data graphically displayed in fig. 8.4. The diagram summarizes the strength of interaction and predicts mechanistic connections among the species. To derive the positions of the species on the diagram, one performs MDS on the distance matrix. The distance between points on the diagram represents the strength of correlation between the two species. About 85% of the MDS object is contained in two dimensions. The two-dimensional (2D) projection of the object (fig. 8.5A) is thus a fairly good representation of the correlation distances among all the species in the network.

We can analyze the MDS diagram in fig. 8.5A as if we had no knowledge of the glycolytic pathway and then compare the results to the known network. First, as expected, the measured concentrations of AMP and citrate in the reactor are tightly correlated to their own input concentrations (AMP-I and citrate-I; see fig. 8.2), and variation in the concentrations of these species is preceded by variation in their inputs (fig. 8.4). Citrate evidently strongly affects the concentration of F26BP: variation in citrate precedes variation in F26BP, and increase of citrate in the system leads to decrease of F26BP. Citrate is positively correlated with F6P and is unlikely to react directly with F26BP;

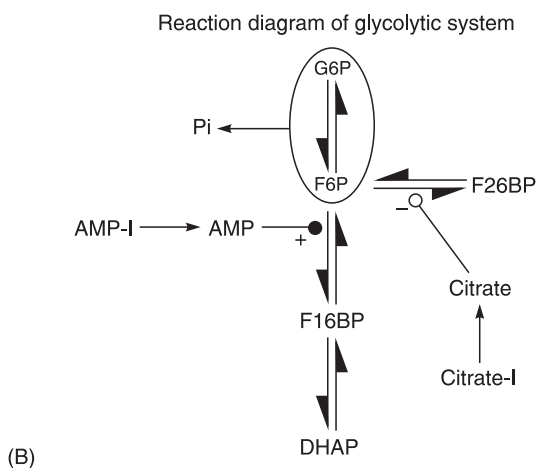
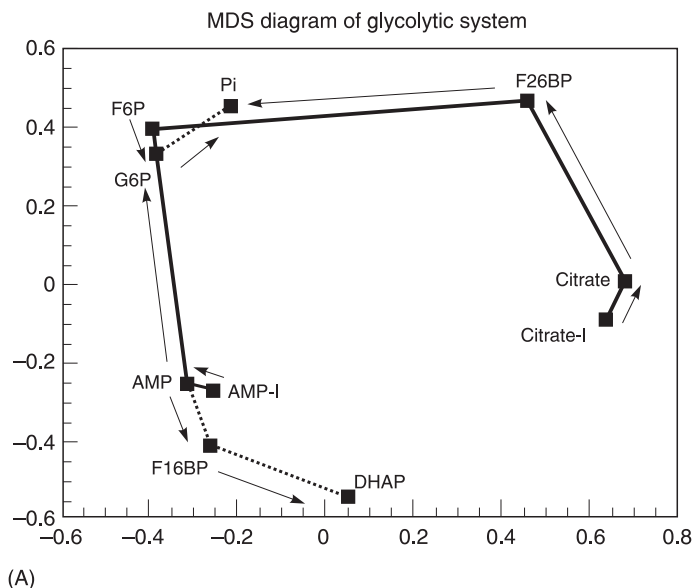


Fig. 8.5 (A) The 2D projection of the MDS diagram for the time series shown in fig. 8.2. Each point represents the time series of a given species. The closer two points are, the higher the correlation between the respective time series. Lines were drawn with the connection algorithm described in chapter 7. Black solid (dotted) lines indicate negative (positive) correlation between the respective species. Arrows indicate temporal ordering among species based on the lagged correlations between their time series. (B) The predicted reaction pathway derived from the CMC diagram. Its correspondence to the known mechanism (fig. 8.1) is high. (From [1].)



it is therefore predicted that citrate either inhibits the conversion of F6P to F26BP or activates the reverse reaction, or both.

The species F6P and G6P are almost too tightly correlated to be distinguished. Given their strong correlation, and because F6P is a simple isomer of G6P, they probably interconvert. The positive correlation between them implies that when material flows into either species, it is quickly repartitioned between them. Therefore, they are likely connected by a quasi-equilibrium reaction. According to temporal ordering, however, a fluctuation in F6P closely precedes the response in G6P; thus, G6P is mechanistically further from the input species than F6P, that is, variation in the inputs must travel through more reaction steps to get to G6P than to get to F6P. Thus, although G6P is connected to AMP by the connection algorithm, consideration of temporal ordering, the strength of the correlation between G6P and F6P, and the chemical argument below imply that the connection from AMP may be to F6P.

The F6P/G6P cluster has a positive correlation with  $P_i$  and negative correlations with F26BP and AMP. This relation implies that F6P and G6P are either converted to these species or are involved in reactions controlled by them. Thus, because it is easy to imagine that F6P is converted to F26BP (and back again), and because phosphate is produced when F6P/G6P is produced from F26BP, the prediction that F6P is interconverted with F26BP is supported. The time ordering of the F26BP–F6P/G6P– $P_i$  part of the diagram may indicate that this reaction pathway favors formation of F6P/G6P under the present conditions.

Chemically, there is no clear route between F6P/G6P and AMP. However, an obvious route exists between F6P and F16BP. The F6P/G6P cluster is negatively correlated with F16BP, implying that the predicted reaction is largely unidirectional, that is, F16BP is formed at the expense of F6P or vice versa. On average, a fluctuation in AMP temporally precedes a response in F16BP and F6P/G6P, but examination of the correlation functions reveals that fluctuations in F16BP closely precede (although they are nearly contemporaneous with) fluctuations in F6P. Because AMP is positively correlated with F16BP and negatively correlated with F6P/G6P, the prediction of the analysis is that AMP activates the conversion of F6P to F16BP (or that AMP inhibits conversion of F16BP to F6P).

Finally, F16BP is strongly positively correlated with DHAP. The positive correlation implies that if the two are chemically interconverted, then the reaction is most likely near a point of quasi-equilibrium. If chemical conversion of F16BP to DHAP is assumed, then there is a three-carbon fragment that we have not measured in this system (this is glyceraldehyde 3-phosphate). Variation in F16BP closely precedes variation in DHAP, implying that DHAP is further, mechanistically, from the inputs than from F16BP.

The pathway as predicted from this analysis (fig. 8.5B) can be compared to the network in fig. 8.1. Both pathways contain the rapid interconversion of F6P and G6P, interconversion of F6P with F26BP and F16BP, and the rapid interconversion of F16BP with DHAP, all of which form the backbone of the partial glycolytic pathway. Citrate is correctly predicted to affect negatively the conversion of F6P to F26BP, and AMP is correctly predicted to activate the conversion of F6P to F16BP.

The choice of specific connections into the F6P/G6P cluster made by the connection algorithm was not expected. One might expect G6P to be less strongly correlated to AMP and F16BP than is F6P, and there is no particular reason why  $P_i$  should be the species most highly correlated with G6P. There are a number of reasons why these assignments were made, but the most important is that the high correlation of G6P with F6P makes

distinguishing between them statistically nearly impossible. Therefore, relatively small amounts of measurement error in the time-series data may lead to reassignment of the connections to F6P and G6P. Inorganic phosphate is most closely associated with the F6P/G6P group because F6P is a product in both reactions that produce the phosphate. Phosphate is produced from the back reactions of F16BP and F26BP to F6P; if you look at phosphate production it would be as correlated with the former two as with the latter.

The inhibition of the conversion of F6P to F16BP by citrate (fig. 8.1) was not predicted. In our experimental system, this inhibition is not present. Most early studies [8,9] of the inhibitory effect of citrate on PFK1 were performed in the absence of F26BP. Later studies have shown that F26BP combined with AMP decreases the inhibitory effect of citrate on PFK1 dramatically [10,11]. In our system, we used an excess amount of PFK2 so that we could measure F26BP by CZE. The concentration of F26BP in the system is about five times higher than its physiological level. This high level of F26BP may be responsible for diminishing the inhibition of PFK1 by citrate.

The CMC theory shows a good prediction of the reaction pathway from the measurements in this much-studied biochemical system. Both the MDS diagram itself (fig. 8.5A) and the predicted reaction pathway (fig. 8.5B) resemble the classically determined reaction pathway (fig. 8.1). In addition, CMC measurements yield information about the underlying kinetics of the network. For example, species connected by small numbers of fast reactions will have smaller distances between them than species connected by a slow reaction. We therefore conclude that the CMC method is useful and predict that it will be applicable to many other complex reaction systems, as is shown in [12] and section 13.5.

There are a number of caveats. The predictions about the reaction pathway provide many essentials of the sequence of steps and types of interactions that exist among the measured species of the network, but do not provide all of the details of the chemical mechanism. Common chemical knowledge, not used yet on purpose, may supply more details. Furthermore, because not all species in the network were measured, we do not predict the entire pathway but only the part specifically involving the measured species. The fact that the calculation yields a reasonable pathway in the absence of full information is a strength of the analysis. Analyses currently under development are concerned with better estimates of the dependence of variation in one variable on variation in the others and detecting the presence of members of the network that are missing from the time series.

The interactions predicted by the CMC analysis are always in reference to the reaction conditions, including pH, temperature, and inflow species concentrations. In our experiment, for example, the inhibitory effect of citrate for conversion of F6P to F16BP was missed, probably because of the nonphysiological concentration of F26BP. If we had seeded the reactor with a different concentration of PFK2::F26BPase, we might have seen this interaction, but we would not have been able to measure F26BP.

Enzymes and effectors on enzymes were not chosen as variables in this example, but could have been so chosen to be measured; the theory still applies.

Even in the face of these difficulties and constraints, CMC provides a powerful method for analyzing multivariate chemical and biochemical time series in order to produce pathway hypotheses and predict likely points of control. Such analyses are becoming increasingly important with the advent of analytical methods for multivariate

time-resolved measurements of chemical and biochemical concentrations such as CZE and parallel gene expression monitoring from DNA chips.

For an application of this method to a genome study, see section 13.5. Other applications of our correlation function method to a part of glycolysis are given in [12] and [13].

*Acknowledgment* This chapter is based in part on the article “A test case of correlation metric construction of a reaction pathway from measurements” by Adam Arkin, Peidong Shen, and John Ross [1], with some changes in wording.

## References

- [1] Arkin, A.; Shen, P.; Ross, J. A test case of correlation metric construction of a reaction pathway from measurements. *Science* **1997**, 277, 1275–1279.
- [2] Stryer, L. *Biochemistry*, 3rd ed.; Freeman: New York, 1988.
- [3] Arkin, A.; Ross, J. Computational functions in biochemical reaction networks. *Biophys. J.* **1994**, 67, 560–578.
- [4] Hocker, C. G.; Epstein, I. R.; Kustin, K.; Tornheim, K. Glycolytic pH oscillations in a flow reactor. *Biophys. Chem.* **1994**, 51, 21–35.
- [5] Oefner, P. J. Surface-charge reversed capillary zone electrophoresis of inorganic and organic anions. *Electrophoresis* **1995**, 16, 46–56.
- [6] Shen, P.; Hauri, D.; Ross, J.; Oefner, P. J. Analysis of glycolysis metabolites by capillary zone electrophoresis with indirect UV detection. *J. Capillary Electrophor.* **1996**, 3, 155–163.
- [7] Card, D. A.; Folmer, D. E.; Sato, S.; Buzza, S. A.; Castleman, A. W., Jr. Covariance mapping of ammonia clusters: evidence of the connectiveness of clusters with coulombic explosion. *J. Phys. Chem.* **1997**, 101, 3417–3423.
- [8] Kemp, R. G. Rabbit liver phosphofructokinase: comparison of some properties with those of muscle phosphofructokinase. *J. Biol. Chem.* **1971**, 246, 245–252.
- [9] Colombo, G. P.; Tate, W.; Girotti, A. W.; Kemp, R. G. Interaction of inhibitors with muscle phosphofructokinase. *J. Biol. Chem.* **1975**, 250, 9404–9412.
- [10] Schaftingen, E. V.; Jett, M. F.; Hue, L. H.; Hers, G. Control of liver 6-phosphofructokinase by fructose 2,6-bisphosphate and other effectors. *Proc. Natl. Acad. Sci. USA* **1981**, 78, 3483–3486.
- [11] Uyeda, K.; Furuya, E.; Luby, L. J. The effect of natural and synthetic D-fructose 2,6-bisphosphate on the regulatory kinetic properties of liver and muscle phosphofructokinases. *J. Biol. Chem.* **1981**, 256, 8394–8399.
- [12] Steuer, R.; Kurths, J.; Fiehn, O.; Weckwerth, W. Observing and interpreting correlations in metabolic networks. *Bioinformatics* **2003**, 19, 1019–1026.
- [13] Kaufmann, K. J.; Pajeroski, J. D.; Jamshidi, N.; Paulson, B. O.; Edwards, J. S. Description and analysis of metabolic connectivity and dynamics in the human blood cell. *Biophys. J.* **2002**, 83, 646–662.

# Density Estimation

---

## 9.1 Entropy Metric Construction (EMC)

In this chapter (based on [1]) we develop further the theory of the correlation method introduced in chapter 7. Consider the expression of the pair correlation function

$$S_{ij}(\tau) = \int (x_i(\tau) - \bar{x}_i)(x_j(t + \tau) - \bar{x}_j) \times p[x_i(t)x_j(t + \tau)] dx_i dx_j \quad (9.1)$$

where an ensemble average replaces the average over a time series of experiments on a single system. The pair correlation function defined in eq. (7.1) is the second moment of the pair distribution function and is obtained from eq. (9.1) by integration.

We choose a new measure of the correlation distance, one based on an information theoretical formulation [2,3]. A natural measure of the correlation distance between two variables is the number of states jointly available to them (the size of the support set) compared to the number of states available to them individually. We therefore require that the measure of the statistical closeness between variables  $X$  and  $Y$  be the fraction of the number of states jointly available to them versus the total possible number of states available to  $X$  and  $Y$  individually. Further, we demand that the measure of the support sets weighs the states according to their probabilities. Thus, two variables are close and the support set is small if the knowledge of one predicts the most likely state of the other, even if there exists simultaneously a substantial number of other states [4–6].

The information entropy gives the distance we demand in these requirements. The effective size of the support set of a continuous variable is [7]

$$S(X) = e^{h(X)} \quad (9.2)$$

in which the entropy  $h(X)$  is defined by

$$h(X) = - \int_S p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x} \quad (9.3)$$

where  $S(X)$  is the support set of  $X$  and  $p(x)$  is the probability density of  $X$ . Similarly, we denote the entropy of a pair of continuous variables  $X, Y$ , as  $h(X, Y)$ , which is related to the pair distribution function  $p(\mathbf{x}, \mathbf{y})$  by an equation analogous to eq. (9.3). The mutual information  $I(X; Y)$  between two stochastic variables is

$$I(X; Y) = h(X) + h(Y) - h(X, Y) \quad (9.4)$$

or in terms of the effective sizes of the support sets,

$$\frac{S(X, Y)}{S(X)S(Y)} = \frac{e^{h(X, Y)}}{e^{h(X)}e^{h(Y)}} = e^{h(X, Y) - h(X) - h(Y)} = e^{-I(X; Y)} \quad (9.5)$$

We define an EMC correlation distance based on information entropy as the minimum of eq. (9.5) regardless of the value of  $\tau$ :

$$d_{xy} = \min_{\tau} \frac{S(X_{\tau}, Y)}{S(X_{\tau})S(Y)} = \min_{\tau} e^{-I(X_{\tau}; Y)} \quad (9.6)$$

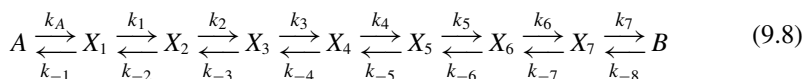
If the correlation of the species  $X$  and  $Y$  is Gaussian, then the expression (9.6) for the EMC distance usually leads to a result similar to that of the CMC distance (eqs. (7.3) and (7.4)).

The CMC and EMC distances satisfy the first three requirements of a metric:

$$\begin{aligned} (1) \quad & d_{xy} \geq 0 \\ (2) \quad & d_{xy} = d_{yx} \\ (3) \quad & d_{xy} = 0 \text{ iff } x \hat{=} y \\ (4) \quad & d_{xz} + d_{zy} \geq d_{xy} \end{aligned} \quad (9.7)$$

but not the triangular inequality, condition (4) in eq. (9.7). This situation can be remedied by the technique of stress minimization [8]. There is frequently little difference in the results of the prediction of a reaction pathway whether the stress minimization technique is applied or not.

It is helpful to treat a specific example of a reaction mechanism to illustrate the calculation of EMC distances and the prediction of a reaction pathway from simulated time series of concentrations. We compare these results with predictions of the CMC method. The example is chosen to show the differences between these two approaches, and to indicate the origin of these differences. The example is the mechanism



All of the reactions are first order with constant rate coefficients, except for the  $A$  to  $X_1$  step, which is enzyme catalyzed. The forward rates of reactions inside the chain are  $k_i = 0.7$  for all  $i$ . The backward rates for the reactions inside the chain are  $k_{-2} = 0.3$ ,  $k_{-5} = 0.2$ , and  $k_{-i} = 0.1$  otherwise. The species  $A$  and  $B$  are held at constant concentrations, with  $A = B = 1$ , and  $X_8$  is varied in a prescribed way (i.e.,  $X_8$  is

chosen to be the input species into the system with unit influx coefficient). Finally, the enzyme-catalyzed step has the effective rate coefficient  $k_A = 60.0X_8/(40.0 + X_8)(60.0 + X_8)$ .

The difficulties in this example arise from the self-inhibition of the enzyme catalysis by  $X_8$ . The rate coefficient  $k_A$  first increases with increasing concentration of  $X_8$  and then decreases. The hypersurface formed by eliminating the time dependence from the set of deterministic rate equations, by dividing the equation for each but one of the species by the equation for that one species, is folded over due to the quadratic dependence of  $k_A$  on  $X_8$ . In the simulation the concentration of  $X_8$  is varied randomly and the responses of the other species are calculated to give time series of 2,000 data points; these series are the starting point for both the EMC and the CMC analysis.

The diamonds in fig. 9.1 show the values of  $X_8$  versus  $X_1$  obtained from these time series. The folding of the hypersurface in the concentration space is shown in this projection onto the  $X_1$ - $X_8$  plane. The space in this plane is divided into rectangles of varying size so that the distribution of points is uniform in each rectangle to within a given accuracy [9]. The density in each rectangle  $(i, j)$ , where  $i$  and  $j$  are indices of the discretization of the continuous ranges of values of  $X_1$  and  $X_8$ , is the pair probability distribution:

$$p_{ij}(x, y) = \frac{N_{ij}}{N_{\text{tot}}A_{ij}} \quad (9.9)$$

where  $N_{ij}$  is the number of points in the particular rectangle labeled  $\{ij\}$ ,  $N_{\text{tot}}$  is the total number of points, and  $A_{ij}$  is the area of the rectangle. We use the pair distribution  $p(X_1, X_8)$  to calculate the singlet distributions  $p(X_1)$  and  $p(X_8)$  by integration. Then, with eq. (9.3), and its analog for  $h(X, Y)$ , we calculate EMC distances (see eq.(9.6)) for

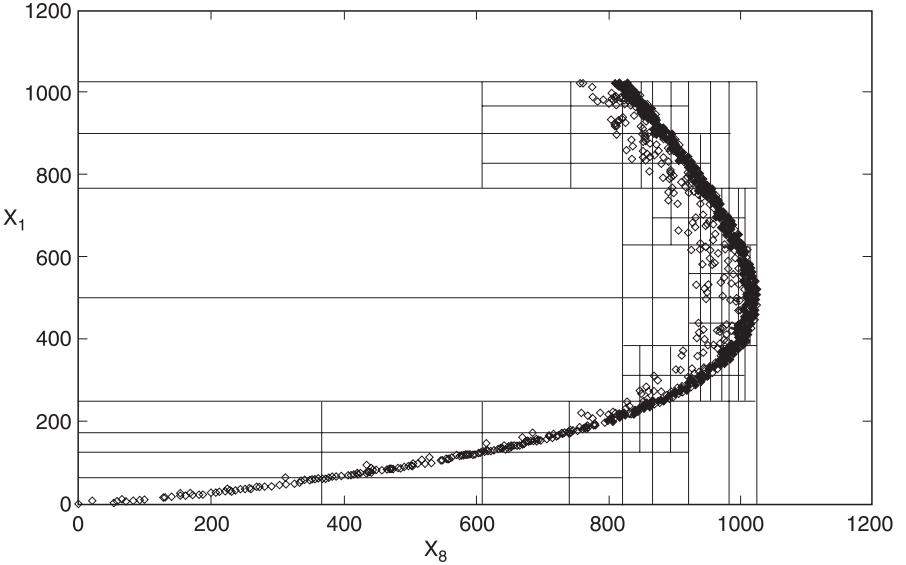


Fig. 9.1 The diamonds plot the values of  $X_1$  versus  $X_8$  obtained from the simulated time series. The rectangles are the result of a partitioning algorithm (see the text). (From [1].)

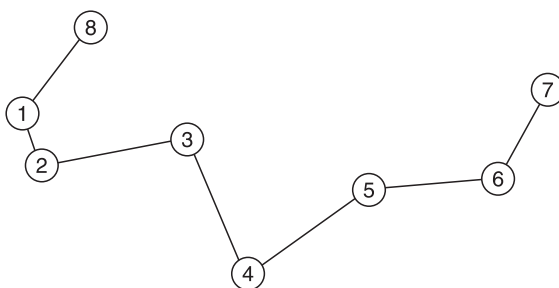


Fig. 9.2 EMC construction of the reaction mechanism of the example. (From [1].)

pairs of species. The primary connections among the species are thus obtained and their corresponding distances, as derived from a multidimensional scaling analysis, are shown in fig. 9.2.

The calculations of the CMC distances by means of eqs. (7.3) and (7.4), and subsequent multidimensional scaling analysis, yields a 2D projection (see fig. 9.3) that differs from the EMC analysis in two respects. The reaction pathway predicted by EMC correctly shows the close correlation of the enzymatic catalyst  $X_8$  and species  $X_1$ . It also shows correctly the chain of linear reactions from  $X_1$  to  $X_7$ . The CMC method also shows that  $X_8$  is correlated with  $X_1$ , but more weakly. Less importantly, the CMC method yields a wrap-around in the placing of species 7. However, the 3D representation of the multidimensional CMC object implemented by stressing the original CMC distances into three rather than two dimensions results in the correct sequencing of the species: the species points lie on a 3D spiral, which when projected onto two dimensions results in the wrap-around effect.

The differences in the CMC and EMC predictions can be traced to the different pair probability densities estimated by these methods from the given time series. In fig. 9.4, we show a contour plot of the pair distribution function  $p(X_1, X_8)$  as calculated by a semi-nonparametric (SNP) method [10] with the time-series simulations for these two species superimposed. In comparison, we show in fig. 9.5 a Gaussian pair probability distribution, as is consistent with CMC, with the same means and variances as those of the EMC distribution. The deviations of the EMC from the Gaussian distribution show that higher than second moments contribute. Since the information entropy for

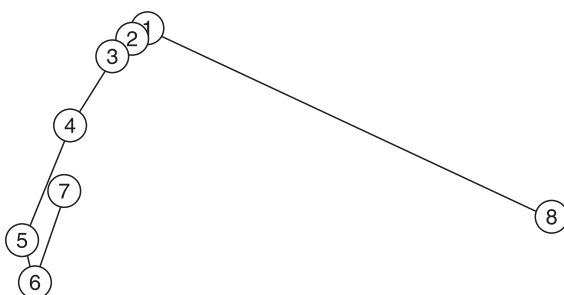


Fig. 9.3 CMC construction of the reaction mechanism given in eq. (9.8). (From [1].)

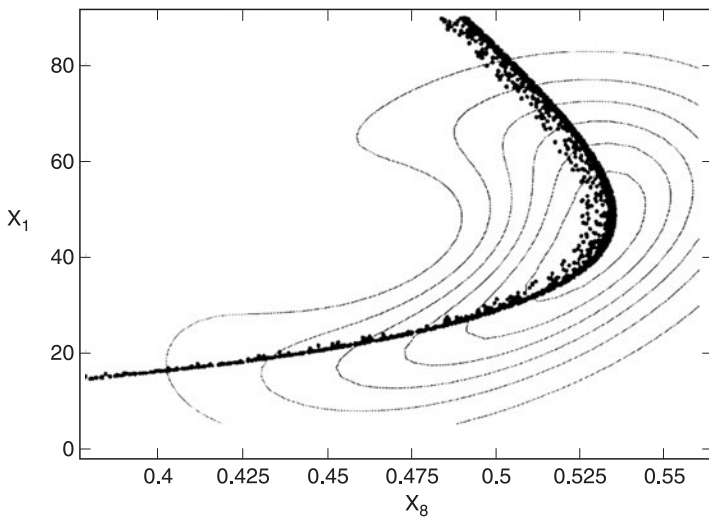


Fig. 9.4 Contour plot of the pair distribution function  $p(X_1, X_8)$ ; the values of the concentrations of  $X_1$  versus  $X_8$ , as obtained from the simulations, are also shown. (From [1].)

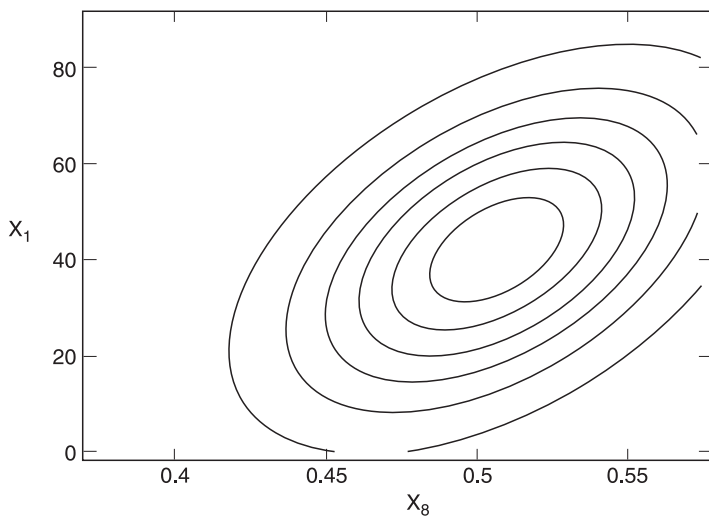


Fig. 9.5 Gaussian joint normal distribution of the variables  $X_1$  and  $X_8$  with the means and covariances of those of the distribution in fig. 9.8. (From [1].)



any distribution is less than or equal to that of a Gaussian distribution, we see that the CMC method gives upper bounds to correlation distances between species.

The CMC method gives a reasonable approximation to the reaction pathways of many systems, including nonlinear systems; strong nonlinearities may demand the use of the EMC method, which requires more data points than the CMC method. We do not know how widespread the need is for the EMC method.

## 9.2 Entropy Reduction Method (ERM)

EMC provides a more accurate representation of the relationship between two variables because nonlinearities are better described if higher moments of the pair distribution are considered. There is, of course, a price to be paid for this higher accuracy and that is the requirement for large amounts of data. However, even with well-estimated pair distribution functions one can, theoretically, make better, less ambiguous predictions of network structure from these measurements.

One possible extension of the EMC method is something we call the entropy reduction method. Here, we essentially require that the nonlinear variation in a variable  $Y$  be explainable by the variations in the concentration of a subset (possibly all) of the other variables in the system. In terms of entropy, this condition is obtained by minimizing the differential entropy  $h(Y|\mathbf{X})$  over the set of variable  $\mathbf{X}$ . The term  $h(Y|\mathbf{X})$  is simply  $h(\mathbf{X},Y) - h(\mathbf{X})$ , where  $h(\mathbf{X},Y) = \int [p(\mathbf{X},Y) \log(p(\mathbf{X},Y)/p(\mathbf{X}))] d\mathbf{X} dY$ . If  $Y$  is completely independent of  $\mathbf{X}$ , then  $h(Y|\mathbf{X}) = h(Y)$ , otherwise it is less than  $h(Y)$ . Formally,  $h(Y|\mathbf{X})$  goes to negative infinity when  $\mathbf{X}$  completely determines  $Y$ . This latter statement is equivalent to saying that the size of the support set,  $\exp[h(Y|\mathbf{X})]$ , is zero. Thus, by iterating through cycles of adding a variable  $x_i$  to  $\mathbf{X}$  that minimizes  $\exp[h(Y|\mathbf{X})]$  until further additions do not decrease the support set, we can find an ordered set of variables that control the variation in  $Y$ .

This technique uses the full joint probability distributions between  $Y$  and the set  $\mathbf{X}$  and so yet again requires a larger amount of data. For multivariate Gaussian distributions, for example, the amount of data needed goes up exponentially with the number of variables [11]. The quality of data and the measurement error, as in the other methods, also play a role in determining the effectiveness of this method. However, it may be shown that this method eliminates a great deal of the ambiguity that arises when only the second moment or a few moments of the distribution are used to define the ordering and causality among the chemical concentration variables. For an expanded discussion of this method see [2].

*Acknowledgment* This chapter is based on selections from the article “On the deduction of chemical reaction pathways from measurements of time series of concentrations” by Michael Samoilov, Adam Arkin, and John Ross [1].

## References

- [1] Samoilov, M.; Arkin, A.; Ross, J. On the deduction of chemical reaction pathways from measurements of time series of concentrations. *Chaos* **2001**, *11*, 108–114.
- [2] Samoilov, M. Reconstruction and functional analysis of general chemical reactions and reaction networks. Ph.D. thesis, Stanford University, 1997.

- [3] Michaels, G. S.; Carr, D.; Askenazi, B. M.; Wen, X.; Fuhrman, S.; Somogyi, R. Cluster analysis and data visualization of large scale gene expression data. *Pacific Symposium on Biocomputing* **1998**, 3, 42–53.
- [4] Liang, S.; Fuhrman, S.; Somogyi, R. REVEAL, a general reverse engineering algorithm for inference of genetic network structures. *Pacific Symposium on Biocomputing* **1998**, 3, 18–29.
- [5] Fuhrman, S.; Cunningham, M. J.; Wen, X.; Zweiger, G. J.; Seilhamer, J.; Somogyi, R. The application of Shannon entropy in the identification of putative drug targets. *BioSystems* **2000**, 55, 5–14.
- [6] Butte, A.; J.; Kohane, I. S. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pacific Symposium on Biocomputing* **2000**, 5, 418–429.
- [7] Cover, T.; Thomas, J. *Elements of Information Theory*; Wiley: New York, 1991.
- [8] Shephard, R. N. Multidimensional scaling, tree-fitting, and clustering. *Science* **1980**, 210, 390–398.
- [9] Fraser, A. M.; Swinney, H. L. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **1986**, 33, 1134–1140.
- [10] Gallant, A.; Tauchen, G. *New Directions in Time Series Analysis, Part II*; Springer-Verlag: New York, 1992.; p 71. *SNP: A Program for Nonparametric Time Series Analysis, Version 8.4. User's Guide*, Chapel Hill, NC, 1995.
- [11] Silverman, B. W. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability 26; Chapman & Hall: New York, 1986; p 94.

# Applications of Genetic Algorithms to the Determination of Reaction Mechanisms

---

The mathematical computational method of genetic algorithms [1] is frequently useful in solving optimization problems in systems with many parameters, for example, a search for suitable parameters of a given problem that achieves a stated purpose. The method searches for these parameters in an efficient parallel way, and has some analogies with evolution. There are other optimization methods available, such as stimulated annealing, but we shall use genetic algorithms.

We shall present three different problems that give an indication of the diversity of applications. We begin with a very short primer on genetic algorithms, which can be omitted if the reader has some knowledge of this subject.

## 10.1 A Short Primer on Genetic Algorithms

Genetic algorithms (GAs) work with a coding of a parameter set, which in the field of chemical kinetics may consist of a number of parameters, such as rate coefficients; variables and constraints, such as concentrations; and other quantities such as chemical species.

Binary coding for a parameter is done as follows. Suppose we have a rate coefficient  $= 9.08 \times 10^{-7}$ ; then if we write that rate coefficient as  $10^{-P}$ , with  $-10 \leq P \leq 10$ , a binary coding with string length of 16 bits is given by

$$P = 10 - 20 \frac{R}{(2^{16} - 1)} \quad (10.1)$$

where  $0 \leq R \leq 2^{16} - 1$ . Since  $P = 6.04$  we have  $R = 12,971$ , or  $R = 0011001010101010$  to the base 2. Thus the value of the rate coefficient is encoded in a single bit string, called a chromosome.

For the solution of a given problem an optimization criterion must be chosen. With a given choice of parameters this criterion is calculated; the comparison of that calculation with the goal set for the criterion gives a fitness value for that set of parameters. If the fitness is adequate but not sufficient, when both are selected by prior choice, for any individual, then retain that individual for the next generation. Reject individuals below that choice.

### 10.1.1 Roulette Wheel Selection

Select individuals for the next generation with a probability proportional to the fitness value from a roulette wheel on which the slot size is proportional to the fitness value. Notice that genetic algorithms use probabilistic, not deterministic, transition rules.

### 10.1.2 Mutation

For the individuals retained from a prior generation we alter their chromosomes. One technique for doing that is mutation. Assign a probability of mutation for any bit in the chromosomes. By selection of random numbers between zero and one for each bit, determine if a mutation is to take place at that bit, from zero to one or vice versa. A mutation takes place at a bit only if the random number chosen for that bit exceeds the probability of the occurrence of a mutation.

### 10.1.3 Crossover

Another technique for creating a new generation from the chromosomes retained from the prior generation is called crossover. We select two individuals from the prior generation, called parents, and let their bit strings be

String 1: 10101101111      String 2: 00000110000

Choose a probability of crossover at each bit and suppose that, according to a selection of random numbers between zero and one, crossover occurs only after the third bit. In the new first string retain the first three bits of the old first string and then add the fourth to the eleventh bit of the old second string. The second new string is formed analogously. Then the children of these parents have chromosomes:

New string 1: 10100110000      New string 2: 00001101111

### 10.1.4 Elitism

If a chromosome attains sufficient fit to the optimization criterion, then retain it for the next generation.

Several other techniques, including variants of the techniques already cited, are available but will not be discussed.

### 10.1.5 Termination

Terminate the optimization when all individuals (chromosomes) have achieved the optimization criterion to the stated sufficiency.

In summary, we create an initial random population of individuals and evaluate their fitness in relation to a chosen optimization criterion. In general the initial fitness is insufficient and a new generation has to be created using the techniques discussed. Evaluate fitness again and keep doing this routine until all individuals meet the optimization criterion.

## 10.2 Selection of Regulation of Flux in a Metabolic Model

We study a simple biochemical reaction model [2] in which certain functional parameters are initially unspecified and are made the object of an optimization procedure. We choose an idealized futile cycle embedded in a larger metabolic pathway (see fig. 10.1). Futile cycles are of interest because they occur widely throughout metabolism and are in some cases implicated as sensitive points of metabolic control. Our model cycle consists of two “irreversible” enzymes (reaction occurs essentially in one direction only) that are regulated by signals from external species, that is, species in the pathway but downstream or upstream from the cycle. We fix the reaction structure of the pathway and certain intrinsic kinetic parameters for the enzymes (the Michaelis–Menten parameters  $K_m$  and  $V_{max}$ ). We further specify that the regulation of the enzymes occurs only by modulation of their  $V_{max}$  through noncompetitive binding of the external species. The parameters governing this modulation, however, are not specified but are to be optimized.

The criterion for optimization is chosen to be flux direction, in which net flux through the cycle is required to be sent in the direction given according to a simple model of biochemical “need.” The cycle carrying out this task idealizes an animal cell that metabolizes blood glucose (F for food in the figure) for energy as long as the glucose concentration in the blood is adequate, but synthesizes glucose for export if the glucose concentration in the blood drops too low. The input or output of a single cell can hardly affect the overall glucose concentration in the blood, so for the cell the blood is a reservoir of glucose. Likewise, the cell’s store of ATP (T for ATP in the figure) may be seen as a reservoir, since various pathways may be active simultaneously in producing

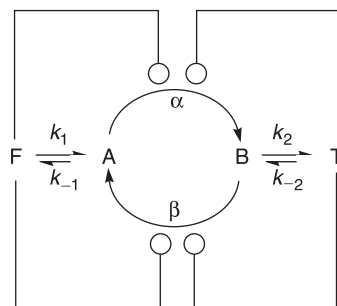


Fig. 10.1 Diagram of the model. F and T are the reservoir species. A and B are the cycle intermediates, interconverted by the enzymes  $\alpha$  and  $\beta$ . Arrows indicate reactions, circles indicate regulation. (From [2].)

and degrading it. The model must funnel material in the correct direction from one reservoir to another in response to concentration levels in those reservoirs, which are under external control, converting its surplus energy to glucose when the glucose level is low, but supplying itself with energy from glucose when its ATP level is low. This task can therefore be seen as a mechanism in the maintenance of homeostasis in both the cell itself and its parent organism. For several previous studies on optimization of metabolic models, not using GA methods, see [3]–[10].

The model, shown in fig. 10.1, consists of three parts: an enzymatic cycle in which a pair of metabolic intermediates, A and B, react reversibly by means of a pair of enzymes,  $\alpha$  and  $\beta$ , and two chemical “reservoirs,” containing metabolic species F and T respectively. These reservoirs represent concentrations of the respective species that at any instant are specified externally. Each of the reservoir species reacts reversibly with one of the cycle intermediates by means of first-order reactions. F is converted into A by a first-order reaction having rate constant  $k_1$ , and A is converted back into F by a reaction having rate constant  $k_{-1}$ . However, because F is a reservoir species, its concentration is unaffected by these reactions, while the concentration of A does change. A similar relationship holds between B and T, the reaction  $B \rightarrow T$  having rate constant  $k_2$ , and  $T \rightarrow B$  having rate constant  $k_{-2}$ . (The kinetic parameters used in the model are given in [2].) The reaction converting A to B is catalyzed by enzyme  $\alpha$  and proceeds with rate  $v_\alpha$ ; the conversion of B to A is catalyzed by enzyme  $\beta$  and proceeds with rate  $v_\beta$ . The kinetic equations describing the temporal variation of A and B are thus

$$\begin{aligned} \dot{A} &= k_1 F + v_\beta - k_{-1} A - v_\alpha \\ \dot{B} &= k_{-2} T + v_\alpha - k_2 B - v_\beta \end{aligned} \quad (10.2)$$

The reservoirs affect the concentration of the cycle species in two ways. The first is through the direct influx represented by the first term in each of eqs. (10.2). The second and more interesting way is through control of the enzyme activities, where the reservoir species F and T are allowed to become effectors of the enzymes. The type of control modeled is noncompetitive allosteric binding of the effectors, where each effector binds to the enzyme independently, as shown in fig. 10.2. In this scheme, the enzyme with effector bound is assumed to have altered catalytic activity toward its substrate compared to that of the enzyme without effector bound. The scheme as shown also relies on the simplifying assumptions that (1) the association and dissociation between enzyme and substrate are unaffected by the binding of the effector, and (2) the binding of substrate to enzyme is much faster than the conversion of bound substrate to product. Under these assumptions, the Michaelis constant  $K_m$  represents the equilibrium constant for

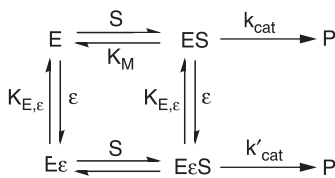


Fig. 10.2 Mechanism of noncompetitive binding used to model the interaction between an enzyme E and an effector  $\epsilon$ . In this approximation,  $K_M$  is the dissociation constant of the enzyme-substrate complex. The binding interaction between the substrate and enzyme is unperturbed by the presence of the effector. Formation of product from the enzyme-substrate complex with effector bound proceeds with the altered rate constant  $k'_{cat}$ . (From [2].)

the dissociation of the enzyme–substrate complex. It is easy to show that the expression for the rate of an enzyme in the presence of effectors is of the form (here shown for enzyme  $\alpha$ ):

$$v_\alpha = \frac{V_{\max,\alpha} A}{K_{M,\alpha} + A} \cdot R_{\alpha,F} \cdot R_{\alpha,T} \quad (10.3)$$

where the form for the factors modifying the intrinsic Michaelis–Menten rate expression is

$$R_{\alpha,\varepsilon} = \frac{K_{\alpha,\varepsilon} + r_{\alpha,\varepsilon} \varepsilon}{K_{\alpha,\varepsilon} + \varepsilon} \quad (10.4)$$

The parameter  $K_{\alpha,\varepsilon}$  is the dissociation constant for the complex of enzyme  $\alpha$  and the effector T or F, labeled  $\varepsilon$ , while  $r_{\alpha,\varepsilon}$  is the ratio of the catalytic rate constants for the enzyme with and without effector bound, respectively. This general form models both activators and inhibitors of the enzyme, depending on whether  $r$  is greater than or less than 1. The response described is hyperbolic, with the half-maximum effect exerted when  $\varepsilon = K$ , and saturating at the maximum effect  $r$  for very high effector concentrations. The cumulative multiplication of modifying factors reflects the assumption that all effectors function at different sites on the enzyme and that the sites do not communicate.

Since the model contains two enzymes, each affected by two reservoir species, there is a total of eight parameters (four  $K$ 's and four  $r$ 's) needed to describe the behavior of the system for given reservoir concentrations. This set of parameters specifies a regulatory pattern which determines the response of the system under differing external concentrations of the reservoirs. The system constitutes a network in which the elements (chemical species) are connected by the described reactions and by regulatory links specified by the regulatory pattern.

The task of flux direction to be carried out by a network is quantified as follows. For each reservoir, a particular concentration value is chosen as optimal. As externally imposed variations decrease the reservoir concentration below this target, a positive state of need is induced. If the reservoir concentration rises above the target, again due to externally imposed variations, a negative need state is induced. The need state is labeled  $\xi_F$  for the reservoir F and  $\xi_T$  for the reservoir T. A functional form for  $\xi$  may be chosen so that there is an acceptable window of concentration around the target, within which the numerical value of the need is close to zero, but at the edges of which the value rapidly changes to a positive or negative one. To resolve conflicts,  $\xi_F$  is taken as a function of both the concentration of F and of the need state of T. In this way, T may override F, an analogy for a cell giving precedence to its own internal state over that of the organism. The particular functional forms for  $\xi_F$  and  $\xi_T$  are specified in fig. 3 of [2]; we do not need them here. The target window for the concentration of T was chosen to be 3 mM to either side of 30 mM, while the target window for F was chosen to be 5 mM to either side of 60 mM.

The need states are used to compute the following function:

$$f \equiv \xi_T(k_2 B - k_{-2} T) + \xi_F(k_{-1} A - k_1 F) \quad (10.5)$$

The term multiplying  $\xi_F$ , the need state for the reservoir F,  $k_{-1} A - k_1 F$ , is the net flux of concentration into that reservoir. Likewise, the term  $k_2 B - k_{-2} T$  that multiplies  $\xi_T$  is the net flux of concentration into the reservoir T. If the net flux into a reservoir has

the same algebraic sign as the reservoir's need state, then there is a positive contribution to  $f$ , while opposite signs result in a negative contribution. A positive overall value of  $f$  thus indicates that the net flux through the cycle is directed properly in accord with the need states, and a negative value indicates that the net flux is directed backwards from what it should be. A zero value of  $f$  results when both reservoirs are within their target concentration ranges. In this case, any value of net flux may be considered proper.

The quantity  $f$  is useful in evaluating the performance of a network under conditions independent of time, such as instantaneous or steady-state conditions. However time-dependent responses are of physiological interest as well. The integral of  $f$  over a period of time  $\tau$ :

$$\mathfrak{F} \equiv \int_0^\tau f dt \quad (10.6)$$

gives some indication of the fraction of the period during which the flux was directed properly. The reservoir concentrations may be changed by external means during the course of such a period, and consequently the need states may change. In order to achieve a high value of  $\mathfrak{F}$ , a network must be able to respond to these changes correctly and quickly.

The measure  $\mathfrak{F}$  may be adjusted to include the effect of an energy "cost" imposed on the operation of the network. Such a cost is incurred when, for instance, ATP is hydrolyzed in an enzymatic reaction. In this study, as an illustration, a cost is imposed on the operation of enzyme  $\alpha$ . The cost is assumed to be expended at the same rate as that of the reaction catalyzed by  $\alpha$ , namely,  $v_\alpha$ . This models a reaction that is driven by the hydrolysis of one stoichiometric equivalent of ATP. The reverse reaction requires no direct metabolic input. The cost in this model is taken to be independent of the concentration of species T, even though T serves as a rough analog of an energy carrier. The overall cost expended over a period of time  $\tau$  is thus the overall flux through  $\alpha$ :

$$C \equiv \int_0^\tau v_\alpha dt \quad (10.7)$$

The cost is weighted and subtracted from  $\mathfrak{F}$  to give a modified (cost-adjusted) measure:

$$\mathfrak{F}' \equiv \mathfrak{F} - mC \quad (10.8)$$

The weighting factor  $m$  used in this study was arbitrarily chosen as 0.001. To achieve a high value of  $\mathfrak{F}'$ , a network must not only be able to respond quickly and correctly to changes in need state, but also simultaneously minimize the energy expended for these responses.

The evaluation of a network's performance of the task proceeds as follows. First, a course of time variation for the concentrations of the reservoirs is constructed. The courses are 3,000 s long, with two concentration regimes, "high" and "low." A sequence of concentration values, typically spaced at 60 s intervals, is then generated so that at each step there is some probability of switching from the current regime (high or low) to the other. The probability of switching between regimes is chosen so that there is on average a switch every 1,200 s, that is, a chance of switching of 1 in 20 per 60 s step.

Second, the eqs. (10.2) are numerically integrated over the course for the network being evaluated, and the results are used to compute  $\mathfrak{F}'$  (eq. (10.8)). Initial conditions for the integration are determined by setting the reservoir concentrations to their optimal



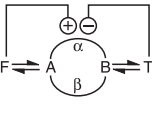
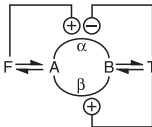
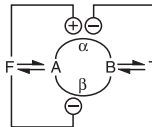
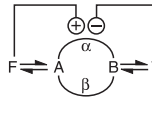
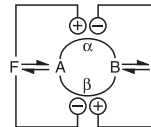
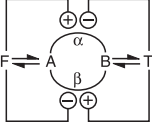
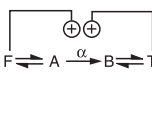
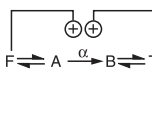
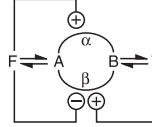
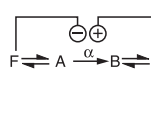
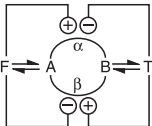
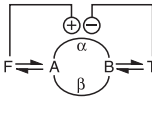
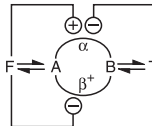
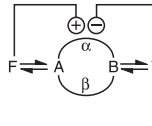
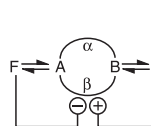
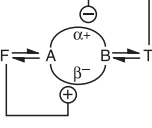
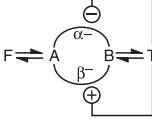
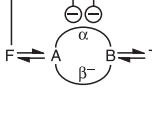
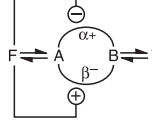
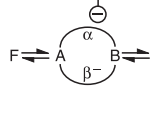
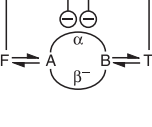
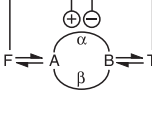
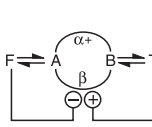
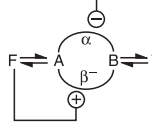
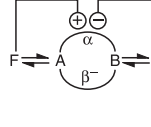
values and letting the network relax to steady state. The higher the value of  $\mathfrak{J}'$  achieved by a network, the better is that network's performance.

To optimize the flux requirement in this model, the set of parameters, the  $K$ 's and  $r$ 's, must be found that maximizes  $\mathfrak{J}'$  in eq. (10.8). This is a difficult optimization for this many parameters and requires a systematic approach, one of a parallel search of the parameter space, as provided by genetic algorithms. The following procedure was used. Networks are encoded as strings ("chromosomes") of eight real numbers ("genes") that are Briggsian logarithms of the  $K$  and  $r$  regulatory parameters. The genetic operators are a two-point cross and a mutation operator. Population size is a constant 100 individuals. The top five and the worst single individual in each generation are reproduced exactly in the next generation. The other 94 slots are filled by recombinant offspring. The parents for each pair of offspring are chosen by roulette wheel selection, with probability proportional to the given parent's share of the combined fitness for the generation (unscaled). The two-point cross is then carried out on the parents by randomly selecting two points along the chromosome. One offspring receives a chromosome that consists of the first parent's genes up to the first cross point, then the second parent's genes up to the second cross point, then the first parent's again until the end of the chromosome. The other offspring receives the complementary set of genes. Each gene of an offspring has a 20% mutation rate, where the mutation operator changes the gene by a random percentage, drawn uniformly from  $-100\%$  to  $100\%$ . A run begins with a population for which the  $\log K$  has a Gaussian distribution with a mean of  $-3.0$  and a variance of  $3.0$ , and the  $\log r$  had a Gaussian distribution with a mean of  $0.0$  and a variance of  $4.0$ . During the run, some individuals fail to be integrated properly (usually owing to the very high stiffness of the resulting kinetic equations) and are replaced by individuals generated randomly according to the above distributions. Most integrator failures occur within the first few generations. The highest fitness for each generation is recorded, and when no improvement in the highest fitness is seen for ten generations, the run is ended.

We now state the results obtained, which in part are given in table 10.1. The five rows of the table correspond to five chosen courses of varying the reservoir concentrations. The GA was run to select the network giving the highest value of  $\mathfrak{J}'$  (eq. (10.8)), determined as described in the previous paragraph. This was repeated five times for each course, and the five resulting networks for each course are shown in the rows of the table in the order they were generated. Networks are designated by a roman numeral indicating the course on which they were generated, followed by an arabic numeral indicating their position in the row. For example, the third network generated on the second course is labeled II.3.

Table 10.1 illustrates regulatory structures of the networks with limiting network diagrams. Such diagrams exaggerate the trends in responsiveness appearing in a network and thus indicate the gross regulatory structure. For example, an effector has no effect on an enzyme if the  $r$  value is 1. In the limiting diagram, if an  $r$  value is close to 1, the connection is not shown. An extremely large value of  $K$  will also cause the connection to be missing in the limiting diagram, because it indicates that the concentration of the effector must become very large for an effect to be exerted. When the  $K$  value is extremely small, the connection is again missing in the limiting diagram, but the enzyme is marked as modified. Since effector concentrations never come very close to zero in our model, an extremely small  $K$  value means that the effector's influence on the enzyme

Table 10.1 Limiting network diagrams of 25 networks grouped by the course on which they were selected

Course					
I					
	Class A	Class A	Class C	Class A	Class A
II					
	Class A	Class D	Class D	Class A	Class D
III					
	Class A	Class A	Class C	Class A	Class A
IV					
	Class G	Class F	Class F	Class C	Class F
V					
	Class E	Class B	Class B	Class E	Class E

Source: From [2].

is saturated for all occurring effector concentrations. The result is the same as if the intrinsic  $V_{\max}$  of the enzyme were itself altered. A plus indicates that the new enzyme has a larger maximum activity than before, and a minus indicates the opposite. The table shows a wide variety of network structures, even on the same course.

The results in table 10.1 lead to several interesting important findings. In all but a few cases, whenever F has an effect on  $\alpha$  it is activation. Similarly, in most cases where T affects  $\alpha$  the effect is inhibitory. For  $\beta$ , the situation is reversed (i.e., F inhibits  $\beta$  while T potentiates it), and the trend is again general. The exceptions are networks II.2, II.3, II.5, IV.1, IV.3, IV.4, V.1, and V.4, but in none of these networks do *all* the connections contradict the trend. Also most of these exceptions perform poorly relative to their siblings (networks generated on the same course), but some nonexceptions do

also. The finding of the trend, that of negative feedback and reciprocal regulation, is important because it is the regulatory pattern expected from a simple, qualitative analysis of homeostasis, which is the task that the system is roughly expected to represent. In a physiological system, the concentration of a reservoir is affected by the fluxes into and out of the reservoir. A homeostatic mechanism should then seek to control these fluxes so that the concentration does not change, at least not significantly. For example, a mechanism to buffer the concentration of  $F$  should inhibit influx into  $F$  as the concentration of  $F$  rises (or potentiate efflux, or both), but do the opposite as the concentration of  $F$  drops. If the signal for these changes is  $F$  itself, and if the reaction structure of the pathway is the one we have modeled, we should expect that  $F$  inhibit  $\beta$ , the enzyme that produces its precursor, and potentiates  $\alpha$ , which consumes its direct product. A similar action for  $T$  is likewise expected. While our model stops short of describing a homeostatic system, in that the reservoir concentrations are not affected by the behavior of the futile cycle, we may imagine reservoir concentrations being changed by the simultaneous action of a large number of identically regulated cycles. In that case, the regulatory pattern we find matches the one expected for homeostasis.

The class designations in table 10.1 refer to different types of switching regions where the concentrations of  $A$  and  $B$  exchange places due to the imposed variations of the reservoir concentrations. In classes  $A$ ,  $B$ ,  $C$ , and  $G$  this switching occurs rapidly, in classes  $E$  and  $F$  slowly, and in class  $D$  not at all. Networks of class  $D$  have eliminated one enzyme. But such gross features do not mark members of the other classes. In fact, for the other classes, the limiting network diagram is not sufficient to determine the class and thus the response behavior. This means that the specific regulatory parameters must be known in addition to the gross regulatory structure. It also means that a class of behavior may be realized by more than one regulatory structure. Further, members of a given class may show rather different performances. For example, class  $C$  is represented by networks I.3, III.3, and IV.4, which, however, have very different performances. Detailed knowledge of regulation is required even for qualitative predictions of performance. A final observation on the results in table 10.1 is that certain classes of behavior are more likely to appear on some courses than others. For instance, course I produced more members of class  $A$  than any other; in fact, only class  $C$  also appeared on that course. On the other hand, course V gave no members of classes  $A$  or  $C$  and gave more members of class  $E$  than  $B$ .

A comparison of performance on the several courses, made in more detail in [2] than presented here, shows that no single network does best on all five courses. There is no single winner as the environment represented by the courses changes, and thus no single species has global dominance. There are some specialists who do particularly well on one course, but of those not all survive the other courses. The winners are *survivors* that perform adequately from course to course, but not necessarily outstandingly; they differ from each other and thus present the opportunity for biological diversity.

### 10.3 Evolutionary Development of Biochemical Oscillatory Reaction Mechanisms

We investigate in this section the application of genetic algorithms to the evolutionary development of a reaction mechanism. We choose as a particular example a part of

glycolysis modeled by the mechanism shown in fig. 10.3 [11]. This reaction can be in a single stable stationary state for given ranges of influx rates of glucose and adenylate pool concentrations (the sum of the concentrations of AMP, ADP, and ATP). However, for different ranges of these two quantities the reaction can be in an oscillatory state, a limit cycle, in which all the metabolite concentrations vary periodically; the variations in this highly nonlinear system are not sinusoidal. (For more on oscillatory reactions, see chapter 11.) Such oscillations have been observed in this reaction, in cell-free extracts of yeast cells, and in whole yeast cells themselves under anaerobic conditions.

Several purposes have been suggested for oscillatory reactions in biological systems. Oscillatory reactions tend to be more stable, and more resistant to external perturbations than comparable systems in a stationary state. The utilization and production of chemical species can be more efficient under oscillatory conditions. For example, in glycolysis with a given, constant input of glucose, the ratio of concentrations of ATP, the energy-rich species, to ADP can be substantially increased in an oscillatory state compared to a stationary state. For an oscillatory input of glucose into an oscillatory reaction mechanism such as fig. 10.3 the ATP/ADP ratio changes; both increases and decreases can be effected with changes in the amplitude and frequency of the oscillatory glucose input. External periodic perturbations of such reactions in concentration, temperature, pressure, light intensity, or imposed electric fields can phase shift the oscillatory rate compared to the oscillatory Gibbs free energy change, with consequent changes in the dissipation and conversely in the efficiency of the reaction [12–16]. Similar phase shifts occur in alternating current (AC) networks, where the analog of the rate is the current, and that of the Gibbs free energy change is proportional to the voltage. These effects of an “AC chemistry” have been shown in experiments on the oscillatory horseradish peroxidase reaction by external periodic variation of the oxygen influx concentration into the system [17,18], in photosynthesis in a C3 plant [19,20], and in theoretical studies in proton transfer [15,21] and combustion reactions [22,23].

Our interest in this problem was kindled by a report on experiments that appeared in 1987 [24]; it showed that the impact of water waves in intertidal zones enhances

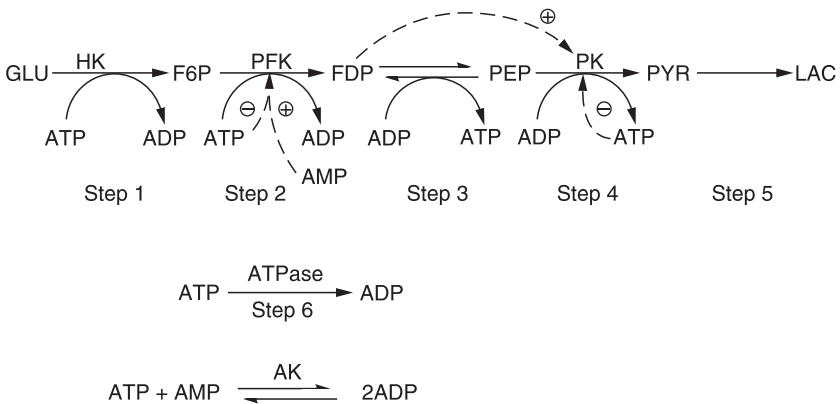


Fig. 10.3 A simple model of the reaction mechanism of glycolysis. For definition of symbols see fig. 6.1. FDP, fructose diphosphate; PYR, pyruvate; LAC, lactate. (From [11].)

Table 10.2 Values of the enzymatic binding constants at autonomous oscillations for constant influx of glucose

Parameter	A	B	C	D	E
	Experimental value	GA (one variable)	$p_{\max}$	GA (two variable)	$p_{\max}$
$\alpha_1 = k_R^{\text{AMP}}/K_{2,T}^{\text{ATP}}$	0.417	0.25–0.45	0.8	0.25–0.45	0.8
$\alpha_2 = K_R^{\text{Fru-1,6-P}_2}/K_{4,T}^{\text{ATP}}$	0.0215	0.075	—	0.053–0.954	0.8
Adenylate pool concentration (mM)	40.95	41.3		41.3	

The values of enzymatic binding constants at autonomous oscillations for constant influx of glucose are given in column A. We define two ratios of binding constants as variables:  $\alpha_1 = K_R^{\text{AMP}}/K_{2,T}^{\text{ATP}}$  for the second reaction (step 2) in fig. 10.3, and  $\alpha_2 = K_R^{\text{Fru-1,6-P}_2}/K_{4,T}^{\text{ATP}}$  for the fourth reaction (step 4). Variable ranges in the GA are given for one-variable and two-variable cases. The maximum change of the parameter  $p_{\max}$  (see text) is determined in such a way that the middle values of variables are in the vicinity of the bifurcation. At the initial values of  $\alpha_1$  and  $\alpha_2$  for both one-variable and two-variable GA cases, this model has a single stationary state. We set the adenylate pool concentration to be 41.3 mM for our study. (From [11].)

the biological growth of intertidal organisms. The authors offer four possible reasons: (1) “waves aid in protecting intertidal inhabitants from their principal enemies”; (2) “waves strip away the boundary layer of used water from kelp blades, thus facilitating nutrient uptake”; (3) “waves apparently enhance algal productivity by allowing algae to use light more efficiently”; and (4) “waves enable some of the shore’s more productive inhabitants to displace their competitors.” Furthermore, they add, “in general, intertidal zones of the northeastern Pacific are more completely covered by plants and animals the more exposed they are to wave action.”

We search for an additional reason, one based on the possibility that oscillatory reactions in biological systems may have developed in evolution in response to periodic perturbations of water waves impacting on rocks at the shore of a sea. For this purpose we use the model reaction mechanism in fig. 10.3. Oscillations in this system may occur and these are due to the presence of feedback and feedforward loops, such as the binding of ATP and ADP to the enzymes PFK and PK, and the binding of FDP to PK. We want to change these binding constants so that no oscillations occur and take that system as our starting point. We abbreviate enzymatic binding and rate coefficients by EBR, and consider four dissociation constants for ligands bound to the  $R$  and  $T$  conformations ( $K_R^{\text{AMP}}, K_{2,T}^{\text{ATP}}, K_R^{\text{AMP}}, K_{4,T}^{\text{ATP}}$ ) for phosphofructokinase (PFKase) and pyruvate kinase (PKase), respectively, and in turn we define two ratios  $\alpha_1 = K_R^{\text{AMP}}/K_{2,T}^{\text{ATP}}$  and  $\alpha_2 = K_R^{\text{Fru-1,6-P}_2}/K_{4,T}^{\text{ATP}}$  as variables. The values of these variables for which autonomous oscillations occur at constant influx of glucose are given in table 10.2; for details on rate coefficients, see [25].

We start the GA procedure by choosing an influx of glucose and the adenylate pool concentration (the sum of the concentrations of AMP, ADP, and ATP) such that this reaction mechanism has a single stable stationary state; in table 10.2, the initial value of  $\alpha_1$  used in the GA is the first number given in the second column and the initial value of  $\alpha_2$  is the first number in the fourth column. We then compare the evolutionary development of two types of such systems, one with a constant influx of glucose and

the other with an oscillatory influx of glucose, to model the periodic effect of waves impinging on an organism. We allow a given initial choice of the EBR to vary by small random increments from one generation to the next, first for one variable and then for two, and carry out this variation systematically by means of a numerical GA method. For our study, we wish to optimize the ratio ATP/ADP, and assume that this ratio is a correlative measure of the rate of growth of an organism. The questions to be asked then are:

1. What is the relative development in time with the GA-controlled variation of the EBR of the two types of systems, one with constant influx of glucose, the other with an oscillatory influx of glucose, with average influx the same as in the constant influx case?
2. How will the ATP/ADP ratio produced by the GA variation of the EBR compare in the two types of systems?
3. What are the changes as we go from one to two variables?
4. What bearings do the answers to these questions have on the issue of the evolutionary development of oscillatory biological reactions and the experimental observations [24]?

The GA works by encoding a parameter of the EBR into a binary string. The algorithm evolves parameter sets by the operations of selection, crossover, and mutations into the strings from one generation to the next. A variable  $p$  has a simple form,  $p_{\text{in}} \cdot (1 + p_{\text{max}} \cdot R / (2^{16} - 1))$ , where  $p_{\text{in}}$  is the initial value of  $p$ ,  $p_{\text{max}}$  gives the maximum change of the parameter not per generation but overall, and  $R$  is initially set to be zero; then a binary string corresponding to  $R$  is evolved within the range  $0 \leq R \leq 2^{16} - 1$  by the GA operations. In the GA, we assign strings for 24 individuals for the variables in the EBR, searching for the change from the initial stable stationary state to autonomous oscillation. We set the maximum change of each variable per generation within  $\pm 7\%$ .

We begin with the genetic algorithm variation of one variable and choose for that the parameter  $\alpha_1$ . Further, we choose the adenylate pool to be constant at 41.3 mM;  $\alpha_2$  is set to be constant at 0.075. The stationary states of the kinetic equations for this system (see [13]) yield the bifurcation diagram shown in fig. 10.4, which is a plot of

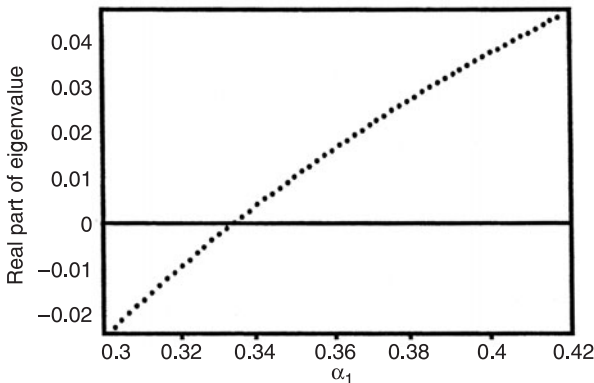


Fig. 10.4 Bifurcation diagram for the adenylate pool at 41.3 mM with  $\alpha_2 = 0.075$ . The real part of the temporal eigenvalues of the linearized kinetic equations is plotted versus the parameter  $\alpha_1$ . The bifurcation occurs after  $\alpha_1 = 0.334$ . (From [11].)

the real part of the temporal eigenvalues of the linearized kinetic equations versus the parameter  $\alpha_1$ . For  $\alpha_1 < 0.334$ , the real part is negative, and the stationary state is either a stable focus or a node. For  $\alpha_1 > 0.334$ , the stationary states are unstable foci and evolve to a stable limit cycle with a period of 0.505 min in the vicinity of the bifurcation. In the application of the GA we compare two different systems; one has a constant influx of glucose (the flow,  $V_{\text{const}} = 2 \text{ mM/min}$ ) and the other has a sinusoidal oscillatory influx of glucose with the same average flux as that of the constant influx, that is,  $V_{\text{osc}} = V_{\text{const}} + a \sin[(2\pi T)t]$ , where  $a$  is the amplitude and  $T$  is the period of the oscillation. For each of the two types of systems, we select 24 copies (individuals), each with an initial value of the parameter  $\alpha_1 = 0.25$ , such that no oscillations occur. We set  $p_{\text{max}}$  to be 0.8, such that the middle value of  $\alpha_1$  (0.35) (see table 10.2) is in the vicinity of the bifurcation. In each new generation, the binary strings of  $\alpha_1$  are evolved by the GA operators within a range such that the new highest value of the parameter in any one generation does not differ by more than 7% from the prior generation. The kinetic equations (see eqs. 1–16 and 21 in [25]) are then solved for each of the 24 individuals and from these solutions we record the state attained for each individual, either stationary or oscillatory. We define the ratio of ATP/ADP in units of its initial value. For the stationary case, the ratio is  $(\text{ATP/ADP})_{\text{stat}}/(\text{ATP/ADP})_{\alpha_1=0.25}$ , where  $(\text{ATP/ADP})_{\text{stat}}$  is the ratio of ATP/ADP in a stationary state and  $(\text{ATP/ADP})_{\alpha_1=0.25}$  is the ratio of the initial stationary state at  $\alpha_1 = 0.25$ . For the oscillatory case, the ratio is  $(\text{ATP/ADP})_{\text{osc}}/(\text{ATP/ADP})_{\alpha_1=0.25}$ , where  $(\text{ATP/ADP})_{\text{osc}}$  is time averaged over one period of the oscillation. In each generation and in each group of 24, we select the system with the highest ATP/ADP ratio and pass that individual to the next generation; we change the worst performer to equal the best performer, and we alter the remaining individuals by mutations and recombinations prior to passing them on to the next generation. As the number of generations increases, the individuals attain higher values of the parameter  $\alpha_1$  up to and beyond the bifurcation.

In the oscillatory influx of glucose, we drive the initially stable focus with a given period (frequency) and amplitude of the imposed sinusoidal glucose oscillation ( $a = 0.5 \text{ mM/min}$ ). Before the bifurcation, the system is either in a stable node or stable focus and responds to the imposed periodic influx of glucose with the same period as that of the influx. After the bifurcation, the period of the response may differ from the period of the autonomous oscillation. Considering the period as a parameter, we observe that if this period is less than the period of the autonomous oscillation (0.5 min) in the vicinity of the bifurcation, the value of the ATP/ADP ratio can be maximized before the bifurcation; for instance, if the driving period is set to be 0.43 min, the value of the ATP/ADP ratio is (locally) maximized before the bifurcation. After the bifurcation, the value of the ATP/ADP ratio decreases with the fixed external period. This decrease is caused by the existence of a narrow entrainment band of the frequency in response to the external oscillatory influx. However, if the fixed external period after the bifurcation is outside the entrainment band [26], then the value of the ATP/ADP ratio is decreased.

The result for the one-variable case is shown in fig. 10.5(a) for an oscillatory influx of glucose. The point at the 30th generation is at the bifurcation. In fig. 10.5(b), the ATP/ADP ratio as a function of  $\alpha_1$  (deterministic) is plotted, and the plot shows bistable behavior after the bifurcation ( $\alpha_1 > 0.334$ ). Note the increase in the ATP/ADP ratio. An analogous plot for a constant input of glucose is given in fig. 10.5(c), in which the

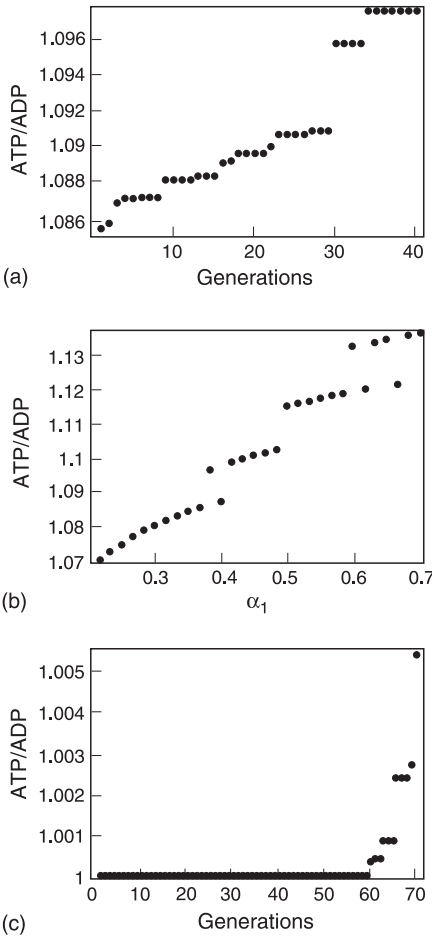


Fig. 10.5 Plot of the ATP/ADP ratio versus (a) the number of generations in the GA, (b)  $\alpha_1$  for the oscillatory influx of glucose with the amplitude 0.5 mM/min (deterministic case), (c) the number of generations for the constant influx of glucose. In (a) and (c), the highest value of the ATP/ADP ratio among 24 individuals at each generation by the GA is plotted and shows that systems with the constant influx of glucose take about double the generations necessary to reach the autonomous oscillation than systems with the oscillatory influx of glucose. (From [11].)

bifurcation is reached at the 60th generation; the value of the ATP/ADP ratio does not change before the bifurcation and only a very small increase in the ATP/ADP ratio occurs just at the bifurcation.

The kinetic equations we solve are deterministic, but stochastic (probabilistic) elements are introduced in the GA. Hence, individuals in a given group arrive at the bifurcation after different numbers of generations. Thus, we generate a frequency distribution for the arrival at the bifurcation as a function of the generation number. This frequency density for an individual reaching the bifurcation differs for the oscillatory input (fig. 10.6a) compared to the constant input of glucose (fig. 10.6b). The most probable value of the generation reaching the bifurcation is 30 for the oscillatory input and substantially larger, 59, for the constant input. The results for the genetic algorithm variation of two variables are similar and are included in the following summary of this study.

The imposition of an oscillatory influx of glucose on the reaction mechanism shown in fig. 10.3, initially in a state of no oscillations (in a node or a focus), may have



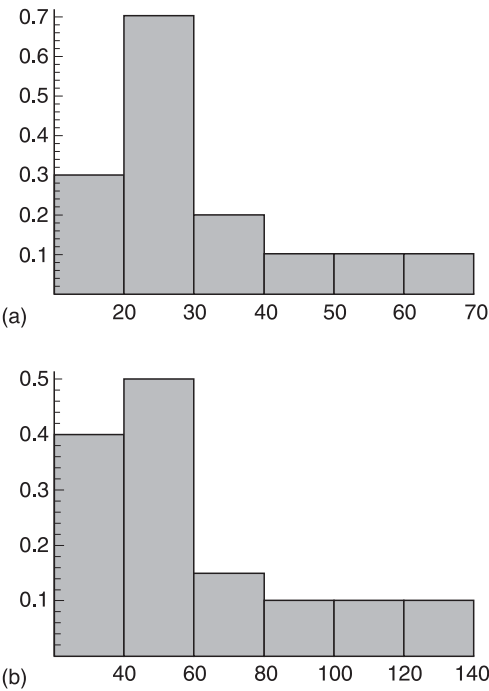


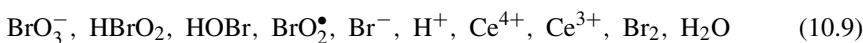
Fig. 10.6 The frequency density versus number of generations to reach the bifurcation: (a) the oscillatory influx (average 30) and (b) the constant influx (average 59). (From [11].)

two evolutionary advantages over the same system with a constant influx of glucose: (1) as shown for the example discussed, a higher ATP/ADP ratio is reached at the bifurcation; and (2) the bifurcation is reached more quickly. For the case of the variation of a single variable with a constant influx of glucose, it takes, for the most probable development, 59 generations to reach oscillatory conditions with increased ATP/ADP ratio of 1.0004. For oscillatory input, this number is reduced to 30 generations with higher ATP/ADP ratio of 1.1. In the variation of two variables, for the constant input case, the number of generations necessary to reach the autonomous oscillation is 167 generations with increased ATP/ADP ratio of 1.0004 and, for the oscillatory glucose input, 76 generations with a higher ATP/ADP ratio of 1.1. We see that the more variables are evolved, the more generations are necessary to reach autonomous oscillations—about tripled (from 59 to 167) for the constant input case and more than doubled (from 30 to 76) for the oscillatory glucose input. These results are obtained from the deterministic kinetic calculation and, with the application of the GA, show the driving force toward the bifurcation.

Both advantages suggest that the evolutionary development of oscillatory reactions, at least in glycolysis, may have occurred at rocks on shores because of the periodic impact of water waves. The faster development of oscillations with periodic wave impact (differences of time scales of many mutations) ensures the spread of that development away from those shores. The individuals subjected to wave impacts have a higher ATP/ADP ratio than those in tidal inlets without such wave impacts, and growth at the shore may therefore expected to be higher, as found by observations [24].

## 10.4 Systematic Determination of Reaction Mechanism and Rate Coefficients

In this section we illustrate the use of genetic algorithms for the determination of reactions mechanisms and rate coefficients from measurements [27]. We apply the optimization method to the cerium catalyzed minimal bromate system, which has been investigated in some detail. This nonlinear system, when open to mass flow, can be in a single stationary state, a node or a focus, or an unstable focus that suggests a stable limit cycle. The return of a system to a stable node, near that node, is given by an exponential decay with a real rate coefficient; the return to a stable focus is given by an exponential decay with a complex rate coefficient that has a negative real part; in that case the system returns to the stable focus with damped oscillations in concentrations. We take it as given, that is, as determined by experiments, that the chemical species in this system are



This reaction has been modeled by several mechanisms; one is the NFT mechanism [28,29], which is an important part of the FKN (Field–Körös–Noyes) mechanism [30] of the oscillatory Belousov–Zhabotinsky reaction [31,32]. Its complex behavior has been studied experimentally in a continuous-flow stirred tank reactor (CSTR) [33]. The rate equations for such an open system are described in a vector form by

$$\frac{d\mathbf{C}}{dt} = \nu \cdot f(\mathbf{C}) + k_0(\mathbf{C} - \mathbf{C}_0) \quad (10.10)$$

where  $f(\mathbf{C})$  is the mass action rates of individual reactions;  $\mathbf{C}$  is a concentration vector for the species;  $\nu$  is a stoichiometric matrix of the species;  $\mathbf{C}_0$  is a feed ( $\text{BrO}_3^-$ ,  $\text{Br}^-$ ,  $\text{H}^+$ ,  $\text{Ce}^{3+}$ ) concentration vector; and  $k_0(\mathbf{C} - \mathbf{C}_0)$  describes the flow terms with flow rate  $k_0$ . In steady states, there are element balance conditions for the reactions [34]:

$$\mathbf{A} \cdot \delta \mathbf{n} = 0 \quad (10.11)$$

where the system formula matrix (or atomic matrix)  $\mathbf{A}$  is given by

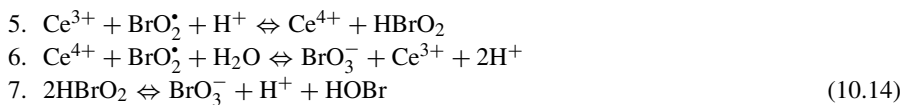
$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 2 & 0 & 0 \\ 3 & 2 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \quad (10.12)$$

and the element balance vector  $\delta \mathbf{n}$  is

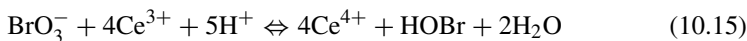
$$\delta \mathbf{n} = \left( \delta_{\text{BrO}_3^-}, \delta_{\text{HBrO}_2}, \delta_{\text{HOBr}}, \delta_{\text{BrO}_2^\bullet}, \delta_{\text{Br}^-}, \delta_{\text{H}^+}, \delta_{\text{Ce}^{4+}}, \delta_{\text{Br}_2}, \delta_{\text{Ce}^{3+}}, \delta_{\text{H}_2\text{O}} \right) \quad (10.13)$$

The set of null-space vectors of the matrix  $\mathbf{A}$  represents the independent reactions that satisfy the stoichiometric condition of mass conservation. If we select all possible bimolecular reactions among all the species, but permit  $\text{H}_2\text{O}$  and  $\text{H}^+$  to be additional reactants, and impose conservation of atoms and charge, then we find that linear combinations of the independent reactions yield seven elementary steps:

1.  $\text{Br}^- + \text{BrO}_3^- + 2\text{H}^+ \rightleftharpoons \text{HBrO}_2 + \text{HOBr}$
2.  $\text{Br}^- + \text{HBrO}_2 + \text{H}^+ \rightleftharpoons 2\text{HOBr}$
3.  $\text{Br}^- + \text{H}^+ + \text{HOBr} \rightleftharpoons \text{Br}_2 + \text{H}_2\text{O}$
4.  $\text{BrO}_3^- + \text{H}^+ + \text{HBrO}_2 \rightleftharpoons 2\text{BrO}_2^\bullet + \text{H}_2\text{O}$



which constitute precisely the NFT mechanism. This result is accidental; for example, it is not the result in the Citri–Epstein mechanism for the chlorite–iodide reaction [35]. In the literature [30] the NFT mechanism is represented by the following overall reaction:



We now check the consistency between the overall reaction (eq. (10.15)) and the elementary steps (eq. (10.14)), we need to find out how many of each of the elementary steps (eq. (10.14)) are necessary for the generation of the overall reaction. We designate the stoichiometric vector attached to the overall reaction (eq. (10.15)) as

$$\mathbf{n}_{\text{overall}} = (-1, 0, 1, 0, 0, -5, 4, 0, -4, 2) \tag{10.16}$$

where the negative sign is for a reactant and the positive for a product; see the order of species in eq. (10.13). From the overall reaction (eq. (10.15)) we see that there are four intermediate species,  $\text{Br}^-$ ,  $\text{HBrO}_2$ ,  $\text{BrO}_2^\bullet$ ,  $\text{Br}_2$ . To obtain the overall reaction, we add up the elementary steps, each of the elementary reactions being multiplied by a stoichiometric number [36]. In the result, we require that the stoichiometric coefficients of the intermediates be equal to zero. In other words, the total number of  $i$ th intermediates created by all  $S$  steps,  $f_i$ , must vanish, that is,

$$f_i = \sum_{s=1}^S b_{is} n_s = 0 \tag{10.17}$$

where  $n_s$  is the stoichiometric number for the  $s$ th reaction step, and  $b_{is}$  is the stoichiometric coefficient of  $i$ th intermediates in the  $s$ th step. Equations (10.17) express the condition that in the overall reactions the stoichiometric coefficients of the active intermediates are equal to zero. Therefore, the null-space of the matrix  $\mathbf{B}$  represents the overall reactions. To satisfy the overall reaction (eq. (10.15)), there must be included at least one of the following vector sets of elementary reaction steps in the NFT mechanism, (eq. (10.14)):

$$\mathbf{v} = \{(4, 5, 7), (4, 6, 7), (5, 6, 7), (1, 2, 4, 5), (1, 2, 4, 6), (1, 2, 5, 6)\} \tag{10.18}$$

Consider the set  $\mathbf{v}_1 = (4, 5, 7)$  as an example, where the matrix  $\mathbf{B}$  is given by the stoichiometric coefficients of the intermediates  $\text{Br}^-$ ,  $\text{HBrO}_2$ ,  $\text{BrO}_2^\bullet$ ,  $\text{Br}_2$  in the steps (4,5,7);

$$\mathbf{B}(\mathbf{v}_1) = \begin{pmatrix} 1 & -1 & 2 \\ -2 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{10.19}$$

Then, the null space of  $\mathbf{B}(\mathbf{v}_1)$  yields a stoichiometric number vector  $\mathbf{n}_{\mathbf{v}_1} = (-2, -4, 1)$  and the overall reaction (eq. (10.15)) is  $\mathbf{n}_{\mathbf{v}_1}$ ,  $\mathbf{v}_1$ . As a result, there are 3 three-step, 13 four-step, 16 five-step, and 7 six-step combinations that satisfy this requirement. We apply GAs to search for oscillatory reaction mechanisms among those sets of reactions.

In the search for an oscillatory reaction mechanism we use as optimization criterion available experimental results. Bar-Eli and Geiseler [28] carried out experiments to

determine the domain of oscillations in the subspace of a range of the initial bromide and bromate concentrations at fixed inflow concentrations of cerium ion and acid. The extent of this experimental domain is estimated from their results given in fig. 3 of [28]. To complicate matters, there are unfortunately several sets of rate coefficients that have been obtained from a variety of investigations; there is reasonable agreement among these sets for some of the rate coefficients and substantial disagreement among others (by several orders of magnitude).

The first GA search was for oscillations in the possible reaction mechanisms given that the GA may vary the influx bromide and bromate concentrations over ranges that include the observations in [28]. For that purpose we used the so-called NFT rate coefficients. Under these conditions, no oscillations were found in any of the three possible three-step mechanisms. The GA reveals that there are two different groups of oscillatory mechanisms in the four-step and five-step sets. The first group of reaction sets of elementary steps consists of the reactions (1,2,4,5), (1,2,4,5,6), and (1,2,4,5,7). For this group, the form of the oscillations is shifted only in phase among these three reaction mechanisms; thus, the four-step (1,2,4,5) is an irreducible mechanism and the elementary steps 6 or 7 are not essential for an oscillatory mechanism.

The second group consists of reactions (1,2,3,4,5) and (1,2,3,4,5,6,7). The two sets involve ten species of which five are independent species. The forms of the oscillations in these two sets are identical; therefore, the five-step (1,2,3,4,5) is an irreducible set and steps 6 and 7 are (again) not essential steps in the seven NFT steps with the FKN rate coefficients. In the mechanisms of both groups I and II, there are five essential species,  $\text{HBrO}_2$ ,  $\text{HOBr}$ ,  $\text{BrO}_2^\bullet$ ,  $\text{Br}^-$ ,  $\text{Ce}_3^+$ , which are necessary to have oscillations in the NFT mechanism. Although the above oscillatory domains are near the experimental one, and the periods and shapes of the oscillations are compatible with the experimental results, both group I and II reaction mechanisms with the FKN rate coefficients do not fit experimental results well.

Next, we focus on the four- and five-step mechanisms in groups I and II. We do a GA search for optimal *rate coefficients* that yield unstable foci at four experimental points in the bromide–bromate concentration plane, and find excellent agreements with the experiments.

The GA reveals that the four-step (1,2,4,5) mechanism exhibits oscillations, but not in the experimental domain. On the other hand, for the five-step mechanism (1,2,3,4,5), the GA finds rate coefficients that give very good agreement with all comparable experiments. The period of waves generated by the five-step mechanism is in excellent agreement with the experiments. The shape of waves is very similar to the experimental waves, but a quantitative comparison of the calculated waves with the experimental was not attempted. Although the four-step mechanism displays oscillation, it fails to reproduce the experimental oscillation domain. The five-step mechanism, which is made up of the four steps in the four-step mechanism plus the additional reaction step 3, manages to reproduce the experimental domain. It follows that the reaction step 3 is essential for agreement with the experiments.

With the use of the five forward rate coefficients found by the GA that yield unstable foci for all four experimental condition, we search again which ones of the possible three, four, and five steps have oscillation domains comparable to the experiments. The GA shows that no reaction mechanism other than those of groups I and II have oscillations in the experimental range. Thus, the oscillatory five-step mechanism consisting of the

elementary reactions (1,2,3,4,5) reproduces the experiments and is to be selected for this system.

## 10.5 Summary

The main part of section 10.4 is the application of genetic algorithms to chemical kinetics for determination of the reaction mechanism and rate coefficients for a complex reaction network. We chose the minimal bromate system and developed a computer code that includes the genetic algorithm search method to automate the determination of the reaction mechanisms and (thermodynamically consistent) rate coefficients. For optimization we chose comparison with available experiments on chemical oscillations in this system. The first part of the code incorporates element balance conditions for reactions in steady states to determine possible elementary reaction steps. Then the Horiuti theory [36] is used to determine the possible mechanisms compatible with the overall reaction of this system. Next, we use a genetic algorithm to search which ones of these possible mechanisms have oscillatory domains in the region determined by experiments. The GA with the FKN rate coefficients revealed two groups of oscillatory mechanisms: group I consists of reactions (1,2,4,5), (1,2,4,5,6), and (1,2,4,5,7) in the NFT steps, and group II consists of reactions (1,2,3,4,5) and (1,2,3,4,5,6,7). Group I has nine species (no  $\text{Br}_2$ ) with four independent species, whereas group II has ten species with five independent species. On the basis of the forms, domains, and periods of the observed oscillations, we find that the four-step (1,2,4,5) mechanism in group I and the five-step (1,2,3,4,5) mechanism in group II are irreducible sets. Therefore, for both group I and II mechanisms, steps 6 and 7 in NFT steps are not essential for the oscillations. Both mechanisms with the FKN rate coefficients, however, fail to fit well the experimental oscillatory domain. Thus, the last part of the GA code was used to determine the range of rate coefficients to fit the experimental results.

With the use of the FKN equilibrium constants, the four-step (1,2,4,5) mechanism exhibits oscillation, but not in the right range. On the other hand, the five-step (1,2,3,4,5) mechanism yields very good agreement with comparable experiments with determination of the range of the five forward rate coefficients. The five steps (1,2,3,4,5) are made up of the four-step mechanism plus step 3, which indicates that step 3 is essential for reproduction of the experimental results.

With the generality of the procedures discussed in this chapter, it seems reasonable to suggest that GA algorithms are useful and promising for the determination of reaction mechanisms and rate coefficients of complex reaction networks.

*Acknowledgments* Section 10.2 is based on parts of the article “Genetic-algorithm selection of a regulatory structure that directs flux in a simple metabolic model” by Alex Gilman and John Ross [2]. The retained sections of that article have been rearranged and edited with minor changes in wording.

Section 10.3 is based on parts of the article “Advantages of external periodic events to the evolution of biochemical oscillatory reactions” by Masa Tsuchiya and John Ross [11]. The retained sections of that article have been rearranged and edited with some further interpretations and some changes in wording.

Section 10.4 is based on parts of the article “Application of genetic algorithm to chemical kinetics: determination of reaction mechanism and rate coefficients for a complex reaction

network” by Masa Tsuchiya and John Ross [27]. The retained sections of that article have been rearranged and edited with some further interpretations and changes in wording.

## References

- [1] Mitchell, M. *An Introduction to Genetic Algorithms*; MIT Press: Cambridge, MA, 1996.
- [2] Gilman, A; Ross, J. Genetic-algorithm selection of a regulatory structure that directs flux in a simple metabolic model. *Biophys. J.* **1995**, 69, 1321–1333.
- [3] Heinrich, R.; Holzhütter, H.-G. Efficiency and design of simple metabolic systems. *Biomed. Biochim. Acta* **1985**, 44, 959–969.
- [4] Heinrich, R.; Holzhütter, H.-G.; Schuster, S. A theoretical approach to the evolution and structural design of enzymatic networks: linear enzymatic chains, branched pathways and glycolysis of erythrocytes. *Bull. Math. Biol.* **1987**, 49, 539–595.
- [5] Melendex-Hevia, E.; Isidoro, A. The game of the pentose phosphate cycle. *J. Theor. Biol.* **1985**, 117, 251–263.
- [6] Bray, D.; Lay, S. Computer simulated evolution of a network of cell-signaling molecules. *Biophys. J.* **1994**, 66, 972–977.
- [7] Savageau, M. A. Optimal design of feedback control by inhibition: steady state considerations. *J. Mol. Evol.* **1974**, 4, 139–156.
- [8] Savageau, M. A. Optimal design of feedback control by inhibition: dynamic considerations. *J. Mol. Evol.* **1975**, 5, 199–222.
- [9] Savageau, M. A. Feedforward inhibition in bio-synthetic pathways: inhibition of the aminoacyl-tRNA synthetase by penultimate product. *J. Theor. Biol.* **1979**, 77, 385–404.
- [10] Savageau, M. A.; Jacknow, G. Feedforward inhibition in bio-synthetic pathways: inhibition of the aminoacyl-tRNA synthetase by intermediates of the pathway. *J. Theor. Biol.* **1979**, 77, 405–425.
- [11] Tsuchiya M.; Ross, J. Advantages of external periodic events to the evolution of biochemical oscillatory reactions. *Proc. Natl. Acad. Sci. USA* **2003**, 100, 9691–9695.
- [12] Richter, P. H.; Ross, J. Oscillations and efficiency in glycolysis. *Biophys. Chem.* **1980**, 12, 285–297.
- [13] Termonia, Y.; Ross, J. Oscillations and control features in glycolysis: numerical analysis of a comprehensive model. *Proc. Natl. Acad. Sci. USA* **1981**, 78, 2952–2956.
- [14] Termonia, Y.; Ross, J. Entrainment and resonance in glycolysis. *Proc. Natl. Acad. Sci. USA* **1982**, 79, 2878–2881.
- [15] Ross, J.; Schell, M. Thermodynamic efficiency in nonlinear biochemical reactions. *Annu. Rev. Biophys. Biophys. Chem.* **1987**, 16, 401–422.
- [16] Astumian, R. D.; Baldwin, R.; Li, R. S.; Ross, J. Quadratic response of a chemical reaction to external oscillations. *J. Chem. Phys.* **1992**, 96, 6536–6542.
- [17] Lazar, J. G.; Ross, J. Changes in mean concentration, phase shifts, and dissipation in a forced oscillatory reaction. *Science* **1990**, 247, 189–192.
- [18] Lazar, J. G.; Ross, J. Experiments on the effects of external periodic variation of constraints on the thermodynamics of an oscillatory system. *J. Chem. Phys.* **1990**, 92, 3579–3589.
- [19] Kocks, P.; Ross, J.; Bjorkman, O. Thermodynamic efficiency and resonance of photosynthesis in a C<sub>3</sub> plant. *J. Phys. Chem.* **1995**, 99, 16483–16489.
- [20] Kocks, P.; Ross, J. Kinetic model for (damped) oscillations of transthylakoid pH in plants. *J. Phys. Chem.* **1995**, 99, 16490–16497.
- [21] Schell, M.; Kundu, K.; Ross, J. Dependence of thermodynamic efficiency of proton pumps on frequency of oscillatory concentration of ATP. *Proc. Natl. Acad. Sci. USA* **1987**, 84, 424–428.
- [22] Escher, C.; Ross, J. Reduction of dissipation in a thermal engine by means of periodic changes of external constraints. *J. Chem. Phys.* **1985**, 82, 2453–2456.
- [23] Schram, H.; Schell, M.; Ross, J. Efficiency of a combustion reaction in power production. *J. Chem. Phys.* **1988**, 88, 2730–2734.
- [24] Leigh, E. G., Jr.; Paine, R. T.; Quinn, J. F.; Suchanek, T. H. Wave energy and intertidal productivity. *Proc. Natl. Acad. Sci. USA* **1987**, 84, 1314–1318.

- [25] Termonia, Y.; Ross, J. Oscillations and control features in glycolysis: numerical analysis of a comprehensive model. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 2952–2956.
- [26] Richter, P.-H.; Rehmus, P.; Ross, J. Control and dissipation in oscillatory chemical engines. *Prog. Theor. Phys.* **1981**, *66*, 385–405.
- [27] Tsuchiya, M.; Ross, J. Application of genetic algorithm to chemical kinetics: determination of reaction mechanism and rate coefficients for a complex reaction network. *J. Phys. Chem. A* **2001**, *105*, 4052–4058.
- [28] Bar-Eli, K.; Geiseler, W. Oscillations in the bromate-bromide-cerous system: the simplest chemical oscillator. *J. Phys. Chem.* **1983**, *87*, 3769–3774.
- [29] Bar-Eli, K.; Field, R. J. Simulation of the minimal bromate( $1^-$ ) continuous flow stirred tank reactor oscillator on the basis of a revised set of rate constants. *J. Phys. Chem.* **1990**, *94*, 3660–3663.
- [30] Noyes, R. M.; Field, R. J.; Thompson, R. C. Mechanism of reaction of bromine(V) with weak one-electron reducing agents. *J. Am. Chem. Soc.* **1971**, *93*, 7315–7316.
- [31] Field, R. J.; Noyes, R. M.; Körös, E. Oscillations in chemical systems. II. Thorough analysis of temporal oscillation in the bromate-cerium-malonic acid system. *J. Am. Chem. Soc.* **1972**, *94*, 8649–8664.
- [32] Field, R. J. Experimental and mechanistic characterization of bromate-ion-driven chemical oscillations and traveling waves in closed systems. In *Oscillations and Traveling Waves in Chemical Systems*; Field, R. J., Burger, M., Eds.; Wiley-Interscience: New York, 1985; pp 55–92.
- [33] (a) Geiseler, W.; Bar-Eli, K. Bistability of the oxidation of cerous ions by bromate in a stirred flow reactor. *J. Phys. Chem.* **1981**, *85*, 908–914. (b) De Kepper, P.; Boissonade, J. From bistability to sustained oscillations in homogeneous chemical systems in flow reactor mode. In *Oscillations and Traveling Waves in Chemical Systems*; Field, R. J., Burger, M., Eds.; Wiley-Interscience: New York, 1985; pp 223–256.
- [34] Smith, W. R.; Missen, R. W. *Chemical Reaction Equilibrium Analysis*; Wiley-Interscience: New York, 1982.
- [35] Citri, O.; Epstein, I. R. Systematic design of chemical oscillators. 42. Dynamic behavior in the chlorite-iodide reaction: a simplified mechanism. *J. Phys. Chem.* **1987**, *91*, 6034–6040.
- [36] Horiuti, J. Theory of reaction rates as based on the stoichiometric number concept. *Ann. N.Y. Acad. Sci.* **1973**, *213*, 5–30.

## Oscillatory Reactions

---

### 11.1 Introduction

Oscillating chemical reactions [1–21] have the distinct property of a periodic or aperiodic oscillatory course of concentrations of reacting chemical species as well as temperature. This behavior is due to an interplay of positive and negative feedback with alternating dominance of these two dynamic effects. For example, an exothermic reaction produces heat that increases temperature, which in turn increases reaction rate and thus produces more heat. Such a thermokinetic effect is thus autocatalytic and represents a positive feedback. When run in a flow-through reactor with a cooling jacket, the autocatalysis is eventually suppressed if the reactant is consumed faster than it is supplied. At the same time, the excess heat is being removed via the jacket, which tends to quench the system. The latter two processes are inhibitory and represent a negative feedback. If the heat removal is slow enough so as not to suppress entirely the autocatalysis, but fast enough for temperature to drop before there is enough reactant available via the feed to restore autocatalysis, then there are oscillations in both temperature and concentration of the reactant. Examples of these thermokinetic oscillations are combustion reactions, which typically take place either in homogeneous gaseous [11,12] or liquid phase [13] or in the presence of a solid catalyst [14], thus representing a heterogeneous reaction system.

Of more interest in the present context are reactions where thermal effects are often negligible, or the system is maintained at constant temperature, as is the case with homogeneous chemical reactions taking place in a thermostated flow-through reactor, as well as biochemical reactions in living cells and organisms. Autocatalysis can easily be realized in isothermal systems, where instead of a heat-producing reaction there will typically be a closed reaction pathway, such that species involved are produced faster by reactions along the pathway than they are consumed by removal reactions.

As an example, let us examine the well-known Belousov–Zhabotinsky (BZ) reaction [15,16] of bromate with malonic acid catalyzed by cerium ions in acidic solution.

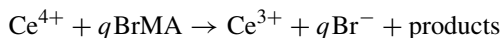


Autocatalysis involves a pathway of two species, the bromic acid  $\text{HBrO}_2$ , and a radical  $\text{BrO}_2^\bullet$ , where the bromic acid reacts with bromate to produce two molecules of the radical, which in turn react with a metal ion catalyst such as  $\text{Ce}^{3+}$  to produce two molecules of bromic acid [17]. So the overall process

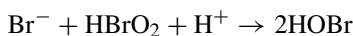


provides net stoichiometric production of bromic acid that makes the autocatalysis possible. More technical conditions for the occurrence of autocatalytic instability are provided by stability analysis of a steady state in association with various theories of reaction networks [10,18]. Later in this chapter we shall use a particular approach called the stoichiometric network analysis [10].

Since the main reactants, bromate and malonic acid, are in surplus, inhibition and negative feedback in the BZ reaction are quite different from that in the previous example of thermokinetic oscillations. Inhibition is provided in part by a direct decomposition of  $\text{HBrO}_2$  (analogous to heat removal in the previous example) but mainly by a chain of reactions of oxidized ion catalyst with brominated malonic acid (BrMA), summarized as



which regenerates the catalyst and provides bromide ions,  $\text{Br}^-$ . The overall stoichiometric coefficient  $q$  is somewhere in the range  $1/2 < q < 1$  depending on external conditions. Bromide ions in turn control the autocatalytic species  $\text{HBrO}_2$ :



Further reactions (omitted here) produce free bromine that takes part in another sequence of reactions that brominate malonic acid.

The pathway from  $\text{Ce}^{4+}$  to  $\text{Br}^-$  is the key to negative feedback on  $\text{HBrO}_2$ , which helps to remove this species after being accumulated due to autocatalysis. However, different roles are associated with the two species: while  $\text{Br}^-$  is an inhibitor that directly removes  $\text{HBrO}_2$ ,  $\text{Ce}^{4+}$  is a controlling species that provides a delay allowing the autocatalysis to advance considerably before inhibition by  $\text{Br}^-$  causes the concentration of  $\text{HBrO}_2$  to drop; hence oscillations may appear. As before, proper time scales of reaction steps are necessary for oscillations to appear. Clearly, this example of isothermal oscillations is more involved than the thermokinetic one. In particular, there are three main (or essential) types of variables rather than two; this observation prompts for a classification of oscillatory reaction mechanisms—one of the main topics of this chapter.

Reactions along the autocatalytic pathway are very fast; typical time scales are fractions of a second. Likewise, the reactions removing the autocatalytic species are rapid, whereas the reaction of a transport process controlling the negative feedback (inflow of the reactant in the thermokinetic example, and conversion of BrMA to  $\text{Br}^-$  in the BZ reaction) is slower by one or more orders of magnitude. This typically produces rather sharp oscillatory peaks and in extreme cases results in a series of spikes interconnected by a recovery phase of relative quiescence; the period of oscillations is typically in the range from tens to hundreds of seconds. The range of concentrations is also varied. The autocatalytic and inhibitory species tend to oscillate within several orders of magnitude in their concentration, whereas the species controlling the negative feedback

typically has a smaller span. Still other species may have yet smaller amplitude and may be buffered with no suppression effect on oscillations (nonessential species).

There are many examples of oscillating chemical reactions provided in the literature [1,19–21]. But the main realm of oscillatory behavior is in biochemical reactions that take place *in vivo* [22–27]. Oscillating biochemical reactions are used by cells as pace-makers [28], means of signaling and information processing [29], metabolic control such as maintaining stable energy supply [30], and output of products [31]. The efficiency of oscillatory reactions is discussed briefly in section 10.3. One example, glycolytic oscillations, has been examined throughout this book. Other examples include repeated firings of neurons, bursting in pancreatic cells, pacemaker cells in the heart, oscillations of calcium ions in cytosol, proton pumps, muscle contraction, and so on (see, for example, Keener and Sneyd [32] and Fall et al. [33]). Among those, intracellular  $[\text{Ca}^{2+}]$  spiking [34,35] is of particular interest, since calcium ions are critically important for many cellular functions including muscle contraction, cardiac electrophysiology, hormone secretion, and others. Oscillations also play an important role in the transcriptional control in genetic networks [36,37].

Much research on oscillatory reactions has been concerned with the complex reaction mechanisms, with many variables and many elementary steps, that give rise to temporal oscillations of concentrations of species in the mechanism. Hypothesized reaction mechanisms have been tested by comparing calculated periods (frequencies), amplitudes, and the shape of temporal variations of oscillating species with experimentally observed behavior. This is an important but frequently not stringent test; also such comparisons are usually not suggestive of improvements in a proposed reaction mechanism. In the last few years, a series of theoretical studies and suggestions of new types of experiments have provided more stringent tests of a proposed mechanism and have led the way to a strategy for formulating reaction mechanisms, based on operational prescriptions, that is, a set of experiments. In this chapter we outline these developments. In section 11.2 we describe several necessary theoretical concepts and constructs of oscillatory reactions. In section 11.3 we list and briefly describe a series of experiments showing what information about the reaction mechanism can be deduced from each type of experiment. We also give examples of the application of these techniques to various experimental systems. In section 11.4 we illustrate the strategy of carrying out the experiments described in section 11.3 on two experimental systems: the horseradish peroxidase oscillatory reaction and the chlorite–iodide reaction.

We focus here on strategies of formulating reaction mechanisms of oscillatory chemical reactions in open (flow-through) and closed (oscillating for a limited time) homogeneous systems with no spatial gradients (either because of stirring or because of relatively fast diffusion in small cell-sized systems).

## 11.2 Concepts and Theoretical Constructs

A homogeneous oscillatory chemical reaction in a closed system is described by a set of ordinary differential equations

$$\frac{dX_i}{dt} = R_i(X_1, \dots, X_n), \quad i = 1, \dots, n \quad (11.1)$$

one for each of the species. Here  $X_i$  is the concentration of species  $i$  and the net reaction rate  $R_i$  for species  $i$  is obtained as a sum over contributing rates from a reaction mechanism consisting of elementary steps by use of mass action kinetics (see section 11.2.4). In an open system (a continuous-flow stirred tank reactor—CSTR) terms are added for inflows and outflows, giving

$$\frac{dX_i}{dt} = F_i(X_1, \dots, X_n) = R_i(X_1, \dots, X_n) + k_0(X_i^0 - X_i) \quad (11.2)$$

where  $k_0(X_i^0 - X_i)$  is the term due to inflow and outflow of the  $i$ th species,  $k_0$  is the reciprocal residence time (or flow rate), and  $X_i^0$  is the input concentration of the  $i$ th species.

### 11.2.1 Jacobian Matrix Elements

The dynamics of the system to first order (i.e., to a linearized approximation) are given by

$$\frac{d\delta X_i}{dt} = \sum_{j=1}^n \left. \frac{\partial F_i}{\partial X_j} \right|_{X^\gamma} \delta X_j, \quad i = 1, \dots, n \quad (11.3)$$

the Taylor expansion of eq. (11.2) about the reference solution  $X^\gamma = (X_1^\gamma(t), \dots, X_n^\gamma(t))$ , where  $\delta X_i = X_i - X_i^\gamma$  is a small perturbation of  $X^\gamma$ . The Jacobian matrix elements (JMEs) are given by the partial derivative  $\partial F_i / \partial X_j = J_{ij}$ . To be explicit, let us rewrite eqs. (11.3) for  $n = 2$ :

$$\begin{pmatrix} \frac{d\delta X_1}{dt} \\ \frac{d\delta X_2}{dt} \end{pmatrix} = \begin{pmatrix} \frac{\partial F_1}{\partial X_1} \delta X_1 + \frac{\partial F_1}{\partial X_2} \delta X_2 \\ \frac{\partial F_2}{\partial X_1} \delta X_1 + \frac{\partial F_2}{\partial X_2} \delta X_2 \end{pmatrix} = \begin{bmatrix} \frac{\partial F_1}{\partial X_1} & \frac{\partial F_1}{\partial X_2} \\ \frac{\partial F_2}{\partial X_1} & \frac{\partial F_2}{\partial X_2} \end{bmatrix} \begin{pmatrix} \delta X_1 \\ \delta X_2 \end{pmatrix} = J \begin{pmatrix} \delta X_1 \\ \delta X_2 \end{pmatrix} \quad (11.4)$$

where  $J = \{J_{ij}\}$  is the Jacobian matrix. When the reference solution chosen is a stationary state, that is,

$$F_i(X^\gamma) = 0 \quad i = 1, \dots, n \quad (11.5)$$

the JMEs give useful information about both the stability of the stationary state and the connectivity among the species in the reaction mechanism of the system. The time evolution of the deviation  $\delta X$  from the stationary state is given in terms of the eigenvalues  $\lambda_k$  and eigenvectors  $w^k = (w_1^k, \dots, w_n^k)$  (where  $k = 1, \dots, n$ ) of the Jacobian matrix,

$$\delta X_i(t) = \sum_{k=1}^n a_k w_i^k \exp^{\lambda_k t}, \quad i = 1, \dots, n \quad (11.6)$$

where the numbers  $a_k$  specify the initial deviation at  $t = 0$ . The stability is determined by the eigenvalues, namely, if real parts of all  $\lambda_k$ 's are negative, then the stationary state is stable.

Tyson [38] classified destabilizing processes in chemical reaction systems according to mathematical relations among the JMEs. He distinguished direct autocatalysis,

which includes product activation and substrate inhibition; indirect autocatalysis, as seen in competition, symbiosis, and positive feedback loops; and negative feedback loops. Luo and Epstein [19] extended Tyson's classification, emphasizing the important interplay between them in oscillating chemical systems. They distinguished three distinct types of negative feedback: coproduct autocontrol, double autocatalysis, and flow control. Chevalier et al. [7] presented new experimental strategies for measuring JMEs at a stationary state of a complex reaction and a discussion of how JMEs may be used to construct a reaction mechanism (see section 11.3.7).

### 11.2.2 Bifurcation Analysis

A given reaction mechanism may have various dynamic regimes displayed in the space of variables (phase space), corresponding to different regions in the space of constraints or parameters. Variables are the quantities  $X_i$  in eqs. (11.2) such as concentrations of reactants, intermediates, and products, and other dynamical quantities, such as temperature in nonisothermal systems. Constraints are imposed, experimentally controllable conditions, such as influx of reactants into an open reaction system, whereas parameters specify properties of the system; examples of parameters are rate coefficients.

The ranges of various dynamic regimes are often displayed in a two-dimensional (2D) bifurcation diagram. An example is given in fig. 11.1, obtained by calculation [5] from a model of the chlorite–iodide reaction [39,40]; the 2D space of constraints in

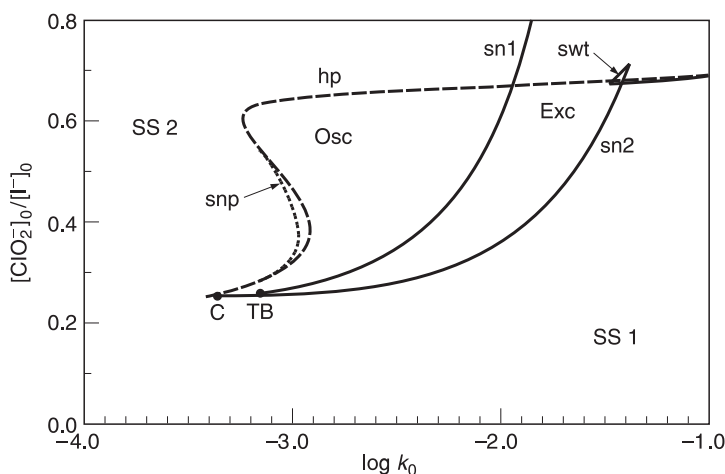


Fig. 11.1 Two-dimensional bifurcation diagram calculated by continuation from the Citri–Epstein mechanism for the chlorite–iodide system. Plot of 2D space of constraints is chosen to be the ratio of input concentrations,  $[\text{ClO}_2]_0/[\text{I}^-]_0$ , versus the logarithm of the reciprocal residence time,  $\log k_0$ . Notation: SS1, SS2, region of steady states with high  $[\text{I}^-]$  and low  $[\text{I}^-]$ , respectively; Osc, region of periodic oscillations; Exc, region of excitability; sn1, sn2, curves of saddle-node bifurcations of steady states; hp, curve of Hopf bifurcations; snp, curve of saddle-node bifurcations of periodic orbits; swt, swallow tail (a small area of tristability); C, cusp point; TB, Takens–Bogdanov point (terminus of hp on sn1). (From [5].)

the diagram is chosen to be the ratio of input concentrations,  $[\text{ClO}_2^-]_0/[\text{I}^-]_0$ , versus the logarithm of the reciprocal residence time,  $\log k_0$ . Several regions with different dynamics are shown: two different stable stationary states; a region of periodic oscillations; a region of two coexisting stable stationary states; and a region where one stable stationary state is excitable (i.e., sensitive to small but superthreshold perturbations that evoke a single large-amplitude peak). In addition, there is a small region of a stable stationary state coexisting with oscillations, and a small region of three coexisting stable stationary states. The various regions are separated by boundaries corresponding to different types of transitions or bifurcations: a Hopf bifurcation, separating stationary states from oscillations; and saddle-node bifurcations of stationary states/periodic oscillations, indicating where a pair of stationary states/periodic orbits merge and disappear.

As a complement, a one-dimensional (1D) bifurcation diagram is a plot of one constraint (or parameter) versus a norm or another characteristic of the displayed dynamic regime. The norm can simply be one of the dynamic variables (time averaged in the case of oscillatory regimes) or a combination of more variables, for example, a Euclidean norm (i.e., the length of the vector  $X = (X_1(t), \dots, X_n(t))$  calculated as  $\sqrt{\sum_i X_i^2}$ ). This diagram adds information on the space of variables at the expense of the space of constraints/parameters, the dimension of which is decreased by one. Figure 11.2 shows a 1D bifurcation diagram calculated for the same model used to generate the 2D diagram in fig. 11.1. Here the constraint  $\log k_0$  is retained, where as  $[\text{ClO}_2^-]_0/[\text{I}^-]_0$  is fixed. The diagram indicates various branches of dynamical regimes: stable steady states as solid lines, unstable steady states as dashed lines, and periodic orbits (or limit cycles) as small dashed lines (this branch is mostly stable, only the part between HP and snp is unstable). From this, ranges of  $k_0$  can be identified over which stationary states and

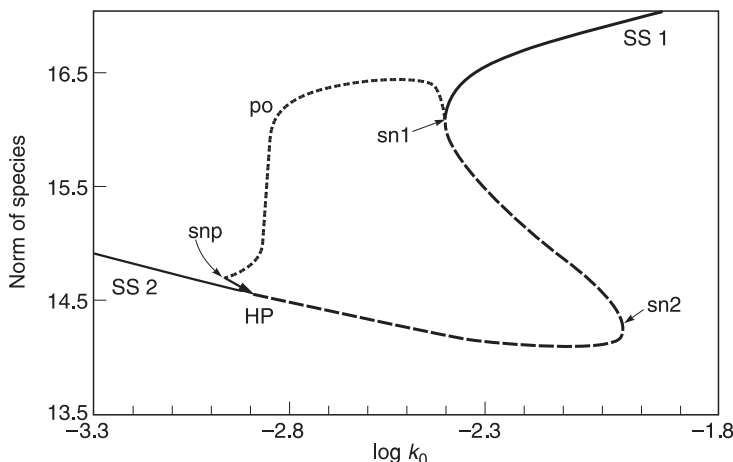


Fig. 11.2 One-dimensional bifurcation diagram calculated by continuation from the Citri–Epstein mechanism for the chlorite–iodide system. The norm for the steady states is  $\sqrt{\sum_i X_i^2}$ , and for the periodic solutions  $\sqrt{1/T \int_0^T \sum_i X(t)^2 dt}$ . Notation: solid lines, stable steady state; dashed lines, unstable steady states; small dashed lines and po, branch of periodic solutions; other notation corresponds to fig. 11.1. (From [5].)

periodic oscillations represent unique stable dynamical regimes, and where the system is multistable. In this case, multistability occurs where the branches of stable SS2 and stable periodic orbits overlap.

Theoretical background for bifurcation analysis can be found, for example, in Guckenheimer and Holmes [41]; calculations of 1D and 2D bifurcation diagrams can be done by continuation methods [42–45]. 2D bifurcation diagrams typically contain a cusp-shaped region of multiple stationary states and essentially delineate regions of oscillatory and nonoscillatory behavior, as well as regions of excitability and regions of bi- or multistability between stationary states or other dynamic regimes. A predominantly occurring structure called the *cross-shaped diagram* [46] involves the cusp-shaped region bounded by a curve corresponding to a saddle-node bifurcation, and a curve corresponding to a Hopf bifurcation, which makes an  $\alpha$ -shaped loop surrounding the cusp point, intersects itself within the cusp-shaped region, and touches the saddle-node curve in two endpoints (Takens–Bogdanov points). Thus the oscillatory region extends beyond the tip of the cusp; bistability of stationary states is confined to the interior of the cusp and excitability occurs in between. This highly symmetric structure (see, e.g., fig. 11.8c) may not always be present; in many cases the self-intersection is missing (as in fig. 11.1) or there are two separate branches of the Hopf bifurcation (as in figs. 11.8d–f).

Predictions of proposed reaction mechanisms are tested against experiments. The details of bifurcation diagrams are useful in seeing the complexity of dynamic regimes of chemical reactions and in providing a series of predictions on dynamic regions and on types of bifurcations to be tested against experiments. Clarke and Jiang [47] provided a method for deriving approximate equations for Hopf and saddle-node bifurcation surfaces of chemical reaction networks, using the methods of stoichiometric network analysis (SNA) to assess the effects of adding or deleting particular reactions on the character of the bifurcation of the network. The methods of SNA will be discussed in more detail in the next section in connection with a categorization of oscillatory reactions.

Some features of bifurcation diagrams are universal, in the sense that if they occur in reaction mechanisms, these features do not distinguish one mechanism from another; other features are specific to a particular reaction mechanism. For example, the increase in amplitude of oscillations of concentrations near a supercritical Hopf bifurcation with a varying constraint is a universal (square root) function. On the other hand, the tilt of the cusp-shaped region in an inflow concentration/flow rate bifurcation diagram is specific to a given species in an oscillatory mechanism, as well as the direction of a transition between high and low stationary state concentration for a given species in a 1D bifurcation diagram. There is an inherent appeal in universality, particularly for mathematicians and physicists. For chemists with their concern for reaction mechanisms, universal features give some general information, such as the presence of a feedback in the reaction mechanism with a Hopf bifurcation; however, specific features may be necessary to distinguish among competing suggested mechanisms.

### 11.2.3 Role of Species in Oscillatory Reactions

Eiswirth et al. [1] proposed a categorization of chemical oscillators. First, a distinction is made between nonessential and essential species. If the concentration of a nonessential species is held constant, oscillations continue. In contrast, if the concentration of an essential species is fixed, oscillations cease. In analyzing the core mechanism of a

chemical oscillator, nonessential species may be neglected or considered as parameters, which simplifies the problem under consideration. The categorization is based on experiments, detailed later, that lead to identification of essential chemical species and their roles in producing the instability leading to oscillatory behavior, and to the connectivity of these species in the elementary steps of the mechanism.

Three types of nonessential species are distinguished. Nonessential species of type A are those that react to produce essential species but do not interact (or only weakly) with other species; type A species can be omitted as dynamic variables and modeled as inflows of the essential species they produce. Nonessential species of type B are inert (or nearly inert) products of the reactions, since, in the limiting case of irreversible reactions, they do not act as reactants anywhere in the network (or do so to a negligible extent); they can thus be left out from the mechanism without (significantly) affecting any other species. Nonessential species of type C react with at least one essential species and may take part in several reactions. Typically, the concentration of type C species may be large and nearly constant; hence buffering the species will have no suppressing effect on oscillations.

In mathematical terms, this classification can be related to the eigenspaces of the Jacobian matrix evaluated at an unstable stationary state surrounded by a stable limit cycle near the supercritical Hopf bifurcation. This bifurcation is characterized by a pair of pure imaginary eigenvalues of the Jacobian matrix. The emerging (small) limit cycle is an ellipse embedded in a plane within the phase space spanned by two eigenvectors that are associated with the pair of bifurcating eigenvalues (center eigenspace). If this plane is perpendicular (or nearly so) to the axis of a particular species, then the limit cycle projects onto this axis as a point (or a small interval), and consequently this species will have (nearly) zero amplitude of oscillations. This is exactly what characterizes type A and type C species. It is convenient to use relative amplitude obtained by scaling the oscillatory amplitude of the species with its stationary value. The remaining  $(n - 2)$  eigenvalues are stable and the corresponding eigenvectors span a stable eigenspace of dimension  $(n - 2)$ ; for three variables, this eigenspace is simply a straight line. If the axis of a species intersects the stable eigenspace at a nonzero angle, then perturbing the oscillations at a specific phase with a specific amount of that species will hit the stable eigenspace and the system will be driven to the stationary state—in other words, the oscillations are quenched [48] (in practice only temporarily, since the stationary state is unstable and ever-present fluctuations will eventually reinstate the oscillations). However, if the axis runs (nearly) parallel with the stable eigenspace, then the size of the quenching perturbation (called the quench vector) would be (nearly) infinite. This is the characteristic of type B and type C species. Therefore relative amplitudes and quench vectors discriminate between all types of nonessential species, and separate them from essential species, as shown in table 11.1. For other ways of determining nonessential species see Stemwedel et al. [2].

The roles of essential species in mechanisms of chemical oscillators are defined as:

1. *Autocatalytic* or *cycle* species (those involved in a reaction in which an increase in the concentration of a species increases the rate of production of that species), denoted X.
2. *Exit* species (those exerting direct inhibitory effect on the cycle species by reacting with it to produce nonessential species or essential species other than X), denoted Y.

Table 11.1 Limiting values of relative amplitudes and quench vectors of nonessential and essential species

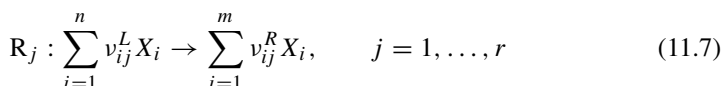
Type	Relative amplitude	Quench vector
A	$\rightarrow 0$	Finite
B	Finite	$\rightarrow \infty$
C	$\rightarrow 0$	$\rightarrow \infty$
Essential	Finite	Finite

3. *Feedback* species (those exerting indirect inhibitory effect on the cycle species by providing negative feedback on the autocatalytic cycle), denoted Z.
4. *Recovery* species (those that take part in reactions allowing the autocatalysis to recover), denoted W.

Not all of these roles are necessary in every category of oscillator.

#### 11.2.4 Stoichiometric Network Analysis (SNA)

In order to discuss the categorization of oscillators set forth by Eiswirth et al. [1], we must first review some of the concepts set out by Clarke [10] in his development of the SNA. Assume that there are  $m$  species taking part in  $r$  chemical reactions so that  $n$  species,  $n \leq m$ , are entering at least one of the reactions:



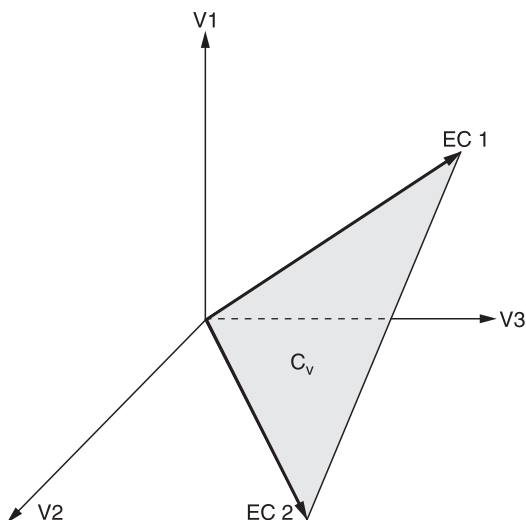
where  $v_{ij}^L \geq 0$  and  $v_{ij}^R \geq 0$  are, respectively, the left-hand and right-hand stoichiometric coefficients of species  $X_i$  in reaction  $R_j$ . The first  $n$  species are assumed to be reactants or intermediates and the remaining  $m - n$  are inert products, which can be left out from further consideration (these are “ideal” type B species). Let  $\bar{S} = \{v_{ij}^R - v_{ij}^L\}$  be the  $(n \times r)$  stoichiometric matrix,  $X = (X_1, \dots, X_n)$  the vector of the chemical species concentrations, and  $\bar{V}(X) = (\bar{V}_1(X), \dots, \bar{V}_r(X))$  the vector of reaction rates. The chemical equations (11.7) together with the vector function  $\bar{V}(X)$  define the mechanism of the reaction, also called a stoichiometric network. Only the concentrations of reactants and intermediates are dynamical variables for which independent dynamical equations can be written. The time evolution of  $X$  in a flow-through system at constant temperature in a well-stirred reaction cell of constant reaction volume can be, according to eqs. (11.2), rewritten in vector notation as

$$\frac{dX}{dt} = F(X) = \bar{S} \bar{V}(X) + k_0(X_0 - X) = SV(X) \quad (11.8)$$

where  $X_0 = (X_{01}, \dots, X_{0n})$  is the vector of feed concentrations for each species and  $k_0$  is the flow rate (or reciprocal residence time). The inflow and outflow terms can be formally treated as zeroth and first-order pseudoreactions, respectively, and are conveniently included in an extended reaction rate vector  $V(X) = (\bar{V}, k_0 X_0, k_0 X)$  having nonnegative components. Accordingly, an extended stoichiometric matrix is  $S = [\bar{S}, I, -I]$ , where  $I$  is an identity matrix. Then  $F(X)$  is simply expressed as  $SV(x)$ .



Fig. 11.3 Three-dimensional reaction rate space spanned by the axes  $V_1$ ,  $V_2$ ,  $V_3$  with a two-dimensional current cone  $C_v$  spanned by two extreme currents  $EC_1$  and  $EC_2$ . The shaded bounded region of the cone is obtained by applying the normalizing condition  $\sum_{k=1}^K \alpha_k E_k \leq 1$ . (From [5].)



A stationary state  $X_s$  satisfies the equation  $SV(X) = 0$ . Hence  $V_s = V(X_s)$  is contained in the null space of  $S$ . Moreover, all components of  $S$  must be nonnegative numbers, which narrows the set of all possible stationary reaction rate vectors  $V_s$  (the currents in SNA terminology) to an open, convex,  $d_r$ -dimensional cone,  $d_r = r - \text{rank}(S)$ , in the space of all  $V$ 's. The edges of this steady-state cone represent sets of steady states that have minimum possible nonzero  $V$ 's admitted by eq. (11.8), and uniquely define a set of major subnetworks (or extreme currents) of the mechanism. In general, the number  $K$  of such subnetworks equals at least the dimension of the cone,  $K \geq d_r$ . An example of a current cone is shown in fig. 11.3.

Let us denote by  $E_k$ ,  $k = 1, \dots, K$  the (arbitrarily normalized) vectors pointing along the edges of the cone. Any linear combination  $\sum_{k=1}^K \alpha_k E_k$  with nonnegative coefficients is again a current. Conversely, any current  $V_s$  can be expressed as a linear combination of extreme currents (such a decomposition is, however, not unique if  $K > d_r$ ). If the  $E_k$ 's are suitably normalized, for example so that  $\sum_{j=1}^r E_{kj} = 1$ ,  $k = 1, \dots, K$ , then the numbers  $\bar{\alpha}_k = \alpha_k / \sum_{k=1}^K \alpha_k$  quantify the contribution of extreme currents to a particular current. Certain subsets of extreme currents span subcones that are  $d$ -dimensional faces of the steady-state cone,  $d = 2, \dots, d_r - 1$ . Hence there is a hierarchy of subnetworks associated with edges and faces of the steady-state cone that may be used as simplified models instead of the full network. The identification of the edges and faces is useful when examining the stability of the (sub)network at a stationary state  $X_s$ . The Jacobian matrix  $J$  of eq. (11.8) at  $X_s$  is

$$J = \left. \frac{dF}{dX} \right|_{X=X_s} = S \left. \frac{dV}{dX} \right|_{X=X_s} = S(\text{diag } V_s) \kappa^T (\text{diag } X_s)^{-1} \quad (11.9)$$

where  $V_s = \sum_{k=1}^K \alpha_k E_k$  and the kinetic matrix  $\kappa = \{\kappa_{ij}\} = \{\partial \ln V_j(X_s) / \partial \ln X_i\}$ . The number  $\kappa_{ij}$  is the effective order of the  $j$ th reaction with respect to the  $i$ th species;

if the reaction rates obey power law kinetics, then  $\kappa_{ij}$  is independent of  $X_s$ . Thus a reparametrization of eq. (11.8) is suggested such that  $X_{s1}, \dots, X_{sn}$  and  $\alpha_1, \dots, \alpha_K$  are new parameters. If power law kinetics is in effect, the stability of the current  $V_s$  is indicated by principal subdeterminants  $\beta_l$  of order  $l = 1, \dots, n$  of the matrix

$$B = -S(\text{diag } V_s)\kappa^T \quad (11.10)$$

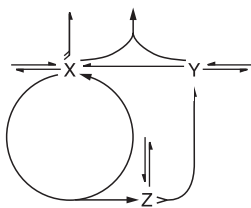
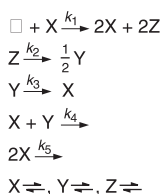
There are  $\binom{n}{l}$  different  $\beta_l$ 's related to all permutations of  $l$  species. If at least one of them is negative, then at least one eigenvalue of  $J$  is unstable provided the values of the steady-state concentrations of the corresponding  $l$  species are sufficiently small [10]. Since  $V_s = \sum_{k=1}^K \alpha_k E_k$ , the stability of the network's steady states depends on the stability of the extreme subnetworks. An unstable  $E_k$  induces instability of the entire network if the corresponding  $\alpha_k$  is large enough and  $X_s$  satisfies the requirement of small concentration of those species for which the corresponding  $\beta_l < 0$ . When linearly combined, the stable  $E_k$ 's usually do not form an unstable current (if they do, then they are called mixing stable), but an instability may occur, since  $V(X)$  is generally nonlinear in  $X$ .

A network diagram is a convenient graphical representation of mass action networks. Any elementary reaction is drawn as a multi-headed multi-tailed arrow oriented from the species entering the reaction to those produced by the reaction: the number of feathers (barbs) at each tail (head) represents the stoichiometric coefficient of the reactant (product); the order of the reaction is the number of left feathers (for examples of network diagrams see fig. 11.4, discussed in the next section). A graph-theoretical approach allows for checking the stability of a (small enough) network by inspection of the graph [10].

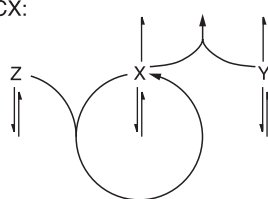
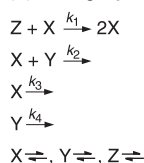
A positive or negative (cyclic) feedback of a species  $i_1$  along a specified cyclic sequence of species  $i_1, i_2, \dots, i_k$  (not necessarily a pathway) states whether an initial (small) perturbation of the stationary state by species  $i_1$  affecting successively the species in the sequence tends to become amplified or damped, respectively. This can be identified by looking at the product of corresponding Jacobian matrix elements,  $J_{i_1, i_k}, J_{i_k, i_{k-1}}, \dots, J_{i_2, i_1}$ . For example, the product  $J_{i_1, i_3} J_{i_3, i_2} J_{i_2, i_1}$  tells us (by reading from the right to the left) what is the change in the stationary value of the species  $i_1$  if an initial (small) perturbation by  $i_1$  affects successively species  $i_2, i_3$ , and finally species  $i_1$  itself again. If the product of JMEs is positive/negative/zero, then the feedback is positive/negative/nonexistent.

If a pathway (i.e., a sequence of species along a causally connected reaction chain, see chapter 1) is cyclic, then we may expect an autocatalytic effect provided there is a positive feedback. However, this condition may be insufficient since other competitive feedbacks may stabilize the dynamics and therefore SNA provides a more subtle way of determining the instability of the stationary state by using the values of  $\beta_l$ . A cyclic pathway is called: (1) a *critical cycle* if the corresponding  $\beta_l$  is equal to zero; (2) a *strong cycle* if the value of  $\beta_l$  is negative; and (3) a *weak cycle* if  $\beta_l$  is positive. The strong cycle is autocatalytic by itself and may provide an oscillatory instability if there is a suitable negative feedback. Then all the cycle species are of type X, and the negative feedback species is of type Z. A critical cycle can be destabilized (and therefore made autocatalytic) on introducing a species that reacts with a cycle species. Such a reaction is called an exit reaction and  $\beta_l$  corresponding to the cycle species combined with the exit species is negative. If there is a suitable negative feedback, then we have again an

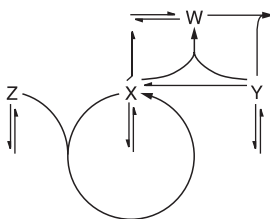
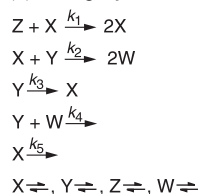
(a) Category 1B:



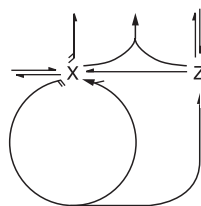
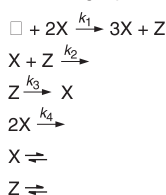
(b) Category 1CX:



(c) Category 1CW:



(d) Category 2B:



(e) Category 2C:

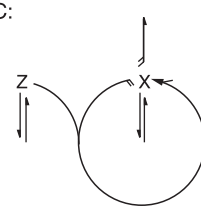
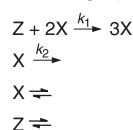


Fig. 11.4 Network diagrams and mechanisms for the prototypes of five categories: (a) category 1B; (b) category 1CX; (c) category 1CW; (d) category 2B; (e) category 2C. (Adapted from [9].)

oscillator and all the cycle species are type X species, the species taking part in the exit reaction are type Y species, and the species providing negative feedback is of type Z. A weak cycle cannot produce instability by adding an exit (or other) reaction. A refined fomulation of detailed conditions for the occurrence of the Hopf bifurcation is provided in [49].

Typically there is a cyclic pathway in oscillatory reaction networks, even though there exist cases when oscillatory instability is associated neither with a cycle in the network [50] nor with a positive feedback [38]. However, the following categorization assumes the presence of a cyclic pathway, since the vast majority of chemical oscillators possess this feature (future extensions are of course possible).

### 11.2.5 Categorization of Oscillatory Reactions

The categories of oscillators discussed in Eiswirth et al. [1] contain topologically different connectivity between the species, that is, the relations among the species in a system, as products and reactants in elementary mechanistic steps, which can produce

oscillatory behavior. Two main categories of oscillators are based on the autocatalytic cycle present in the mechanism. Category 1 is the group of oscillators with a critical cycle combined with an exit reaction; this implies that type Y species is always involved, in addition to type X and Z. Category 2 is a group of oscillators with strong cycles; here a direct removal of a cyclic species instead of an exit reaction is sufficient, and thus type Y is absent. Further classification of mechanisms leads to subcategories 1B, 1C (further subdivided into 1CX and 1CW) and 2B, 2C.

Figure 11.4a–e introduces prototypical examples of these categories. The figure provides elementary mechanistic steps for each prototype along with the corresponding network diagrams, each of them having distinctive topological features as described below.

Subcategories within this classification scheme are distinguished according to the kind of negative feedback involving type Z species:

1. *Input negative feedback* ( $J_{zx}J_{xz} < 0$ ). Type Z species is flowed into the reactor and controls the autocatalytic process by reacting with the type X species; this feature may occur in both category 1 (fig. 11.4b,c) and 2 (fig. 11.4e). Additionally, in category 1, type Z species may also be produced by an exit reaction (i.e.,  $X + Y \rightarrow Z$ ) rather than provided via feed.
2. *Output negative feedback* ( $J_{zx}J_{xy}J_{yz} < 0$ ). Type Z species is a product of a cycle reaction (such a reaction is termed tangent) and controls the autocatalysis indirectly by producing type Y species that inhibits the type X species (tangent feedback, see fig. 11.4a); in another variant of this kind of indirect control the type Z species is produced by exit rather than tangent reaction, and then produces type Y species (exit feedback). For category 2, the feedback expressed by JMEs is formally the same as for the input feedback, but Z is produced rather than consumed by an autocatalytic cycle reaction; in the next step Z reacts directly with X in an exit reaction thereby leaving out type Y species. This represents tangent feedback (fig. 11.4d), but exit feedback may occur as well (e.g., via  $Z + X \rightarrow 2Z$ ).

The subcategories 1C (fig. 11.4b,c) and 2C (fig. 11.4e) involve mechanisms with input feedback; either Z or Y (or both) has to be provided by feed and thus the C subcategory is crucially dependent on inflow (C stands for continuous). The subcategories 1B (fig. 11.4a) and 2B (fig. 11.4d) involve mechanisms with output feedback, tangent or exit. Here the autocatalytic process is driven by a nonessential species (of type A or C) present in surplus and thus all essential species are generated by reactions rather than inflow. Consequently, reactions in the B subcategory provide sustained oscillations even in a closed system until the nonessential species is exhausted (B stands for batch). The 1C subcategory can be further divided into the subcategories 1CW (fig. 11.4c) and 1CX (fig. 11.4b). The former involves a new essential species of type W, which removes the type Y species and thus enables the autocatalysis to recover, whereas the latter accomplishes the same without the presence of W just by a slow inflow of X.

In category 1, the order of the autocatalytic species in the cycle reaction is equal to one (and equal to the order of X in the exit reaction). Category 2 oscillators include mechanisms with autocatalytic reaction of a higher (effective) order in X than the order of X in the removal reaction, and the type X species need not necessarily be chemical species. Rather, they may include vacant surface sites in heterogeneous catalysis, or temperature in thermokinetic oscillation.

### 11.3 Experiments Leading to Information about the Oscillatory Reaction Mechanism

In this section we briefly describe, with examples, a series of experiments from which we can deduce the essential and nonessential species in a given oscillatory system, the connectivity of the essential species producing the oscillations, and the category of the oscillatory system; these steps lead to a formulation of the core of the reaction mechanism. We also describe a variety of experiments for testing a proposed mechanism.

#### 11.3.1 Characterization of Oscillations of Chemical Species

Experiments on the temporal variation of oscillatory species allow a comparison of the amplitude, period, and shape of concentration oscillations observed experimentally with those predicted by a proposed model. Refinements are made in the model to give better qualitative and quantitative agreement.

An example of such a comparison is seen in the modeling of the oscillating chlorite–iodide reaction. The model initially proposed by Epstein and Kustin [39] showed only fair agreement with the experimentally observed  $I^-$  evolution, and worse agreement with the experimentally observed  $I_2$  evolution, as seen in fig. 11.5(a,b). A revised mechanism proposed by Citri and Epstein [40] predicts oscillations quite similar in shape to the experimentally observed  $I^-$  and  $I_2$  oscillations (fig. 11.5c). In many oscillatory systems the temporal variation of only a few species (essential or nonessential) can be measured. The comparison of an experimental time series with a prediction of a proposed mechanism can be made with regard to the period of the oscillations, but becomes subjective with regard to the shape of the variation. The comparisons do not easily lead to suggestions for improvements of the proposed reaction mechanism.

#### 11.3.2 Amplitude Relations

The measurement of the amplitudes of as many oscillating species as possible leads to useful results. An examination of the relative amplitudes allows the identification of essential and nonessential species. Figure 11.6 shows traces of the oscillations of NADH and oxygen in the peroxidase–oxidase (PO) reaction [6]. The relative amplitude of the oscillations in  $O_2$  is seen to be an order of magnitude greater than the amplitude of the oscillations in NADH. Small relative amplitude distinguishes nonessential species of types A and C from nonessential species of type B. By referring to table 11.1, the data presented above suggest that  $O_2$  is essential to oscillations of the horseradish peroxidase system, or is nonessential of type B; however, since  $O_2$  is flowed into the system at a constant rate (i.e.,  $O_2$  is a reactant), we conclude that  $O_2$  is essential. In contrast, the small relative amplitude of NADH oscillations indicates that NADH is a nonessential species of type A or C.

Such a comparison is only useful for analyzing experimental data if oscillations in more than one species can be measured simultaneously or correlated. A comparison of relative amplitudes predicted by a proposed model is probably the most straightforward way to predict nonessential species.

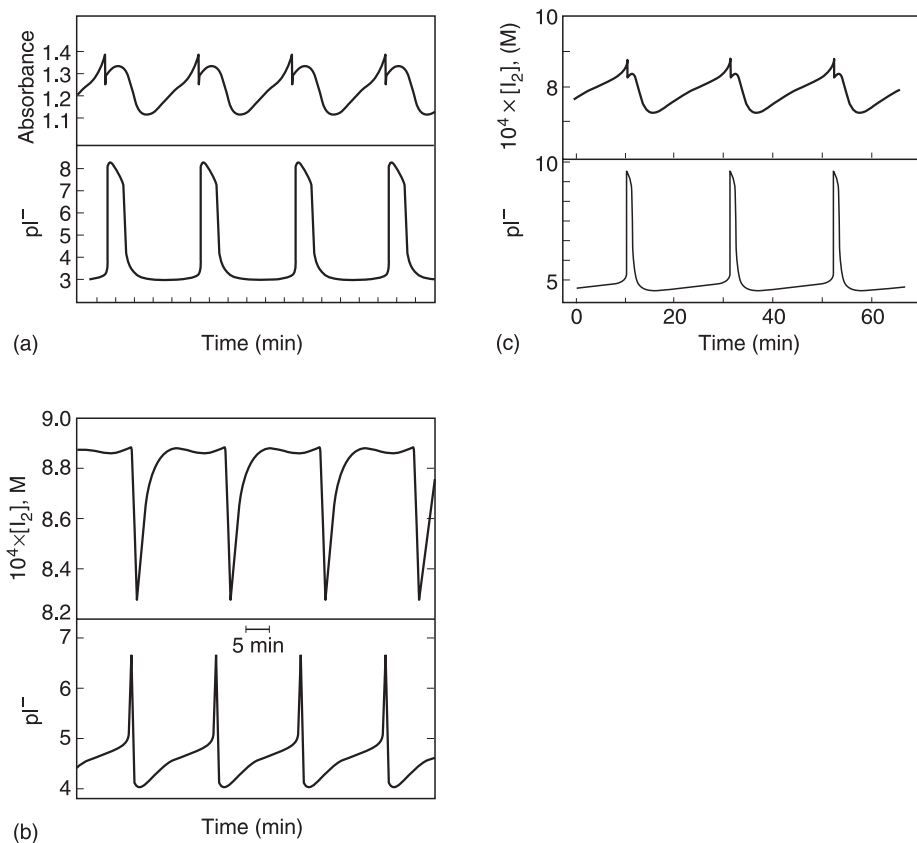


Fig. 11.5 Comparison of experimental and theoretical oscillatory traces of  $I_2$  absorbance and  $I^-$  potential for the chlorite-iodide reaction. (a) Experimental traces. (b) Theoretical traces from the Epstein-Kustin mechanism show fair qualitative agreement, but wave shapes are substantially different from experimental observations. (c) Theoretical traces from the Citri-Epstein mechanism show marked improvement in wave shape agreement with experiments. (From [2].)

### 11.3.3 Phase Relations

From measurements of the temporal variations of chemical species, comparison is made of the phases of as many essential oscillating species as possible. The phases of the species are assigned as in phase (I), antiphase (A), advanced (+), or delayed (−) with respect to the oscillations of a reference species. This results in a sign-symbolic phase shift matrix. The sign-symbolic shift matrix  $\Delta\Phi_{\text{symp}}$  for the prototype of each category is given in table 11.2.

Let us illustrate this technique with the results of simulated measurements of the peroxidase-oxidase reaction based on a skeleton model proposed by Degn et al. [51] (the DOP model). There are four species involved in the model;  $S_1 = O_2$ ,  $S_2 = \text{NADH}$ ,  $S_3$  and  $S_4$  are intermediates associated according to Olsen et al. [52] with free radical

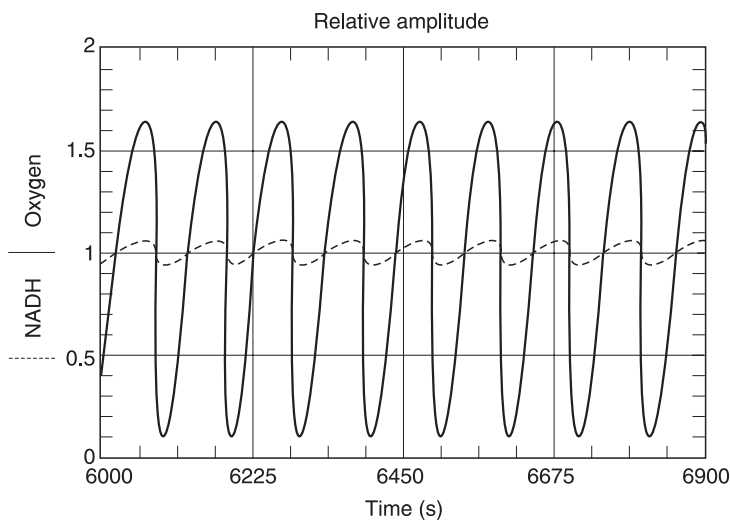


Fig. 11.6 Experimentally determined relative amplitudes of  $O_2$  and NADH in the PO system. Solid line,  $O_2$  concentration; dashed line, NADH concentration; concentrations are taken relative to average values of respective species. (From [6].)

Table 11.2 Sign-symbolic phase shift matrices for the essential species in the prototype of each category of oscillatory reactions (type W in category 1CW behaves as type X)

Category 1B		Phase relation of		
		X	Y	Z
With respect to	X	I	—	—
	Y	+	I	+
	Z	+	—	I

Category 1C		Phase relation of		
		X	Y	Z
With respect to	X	I	A	+
	Y	A	I	—
	Z	—	+	I

Category 2B		Phase relation of	
		X	Z
With respect to	X	I	—
	Z	+	I

Category 2C		Phase relation of	
		X	Z
With respect to	X	I	+
	Z	—	I

Table 11.3 Sign-symbolic phase shift matrix obtained from the DOP model of the peroxidase–oxidase reaction

DOP model		Phase relation of			
		$S_1$	$S_2$	$S_3$	$S_4$
With respect to	$S_1$	I	I	—	—
	$S_2$	I	I	—	—
	$S_3$	+	+	I	I
	$S_4$	+	+	I	I

species  $\text{NAD}^\bullet$  and  $\text{O}_2^{\bullet-}$ , respectively. Chevalier et al. [7] found the sign-symbolic phase shift matrix given in table 11.3.

The entries of this matrix indicate that the oscillations in species  $S_1$  and  $S_2$  are mutually in phase, and likewise with the pair  $S_3$  and  $S_4$ . However, phase relations alone cannot be used to distinguish essential and nonessential species. Chevalier et al. [7] used amplitude relations (see table 11.1) to determine that  $S_1$  was in fact nonessential for particular sets of constraints and then  $S_2$  was of type Z, while for other sets the situation was reversed depending on the choice of external constraints. Let us assume in agreement with experimental conditions that  $S_2$  is nonessential (i.e., NADH is in surplus). The essential species  $S_3$  and  $S_4$  are delayed by  $S_1$  (and  $S_2$ ). Both the rows and the columns for  $S_3$  and  $S_4$  are the same and these two species should play the same role (except possibly one being type X and the other type W, which cannot be distinguished by this method). There is no antiphase relationship in table 11.3, so the category 1C is excluded. Category 1B implies two different out-of-phase species which is not the present case, therefore we conclude that there is no type Y species and so the category is either 2B or 2C, and the species  $S_1$  is either of type X and then  $S_3$  and  $S_4$  are of type Z, or vice versa.

A clue as to which of the two cases is correct cannot be determined based on phase shifts solely and other methods discussed further below should be used. However, there is a considerable asymmetry between both types of species. As a rule, stationary values of Z are at least several times higher than those of X (in the same scale), which helps to make a correct assignment of the role. Taking this into account, species  $S_3$  and  $S_4$  are likely to be of type X and  $S_1$  of type Z. Thus the category is likely to be 2C, which is corroborated by observing that this model requires inflow to provide oscillations.

#### 11.3.4 Concentration Shift Regulation

In this experiment, a constant inflow of one species at a time is added to the system at steady state near a supercritical Hopf bifurcation. This inflow additional to the inflow terms in eq. (11.2) should not be large enough to shift the system from one dynamics regime to another, for example, from a stationary state to an oscillatory state. The response of the concentrations of as many species as possible should be determined after the addition of each species and compared to the steady-state concentrations of the unperturbed system. These measurements allow the construction of an experimental shift matrix, which is directly related to the Jacobian matrix. If we approximate the



Table 11.4 Sign-symbolic concentration shift matrices for the essential species in the prototype of each category of oscillatory reactions (type W in category 1CW behaves as type X)

Category 1B		Perturbing species		
		X	Y	Z
Measured species	X	+	−	−
	Y	−	+	+
	Z	+	−	−

Category 1C		Perturbing species		
		X	Y	Z
Measured species	X	+	−	+
	Y	−	+	−
	Z	−	+	−

Category 2B		Perturbing species	
		X	Z
Measured species	X	+	−
	Z	+	−

Category 2C		Perturbing species	
		X	Z
Measured species	X	+	+
	Z	−	−

change in steady-state concentration  $X_{si}$  with respect to change in inflow concentration  $X_{0j}$  for all  $i$ 's and  $j$ 's as the matrix  $dX_s/dX_0 = \{\partial X_{si}/\partial X_{0j}\}$ , the shift matrix  $\Delta X$  is then given by

$$\Delta X = \frac{dX_s}{dX_0} = -k_0 \left( \frac{dF}{dX} \right)^{-1} \Big|_{X=X_s}$$

(11.11)

where, as in eq. (11.4),  $dF/dX = \{\partial F_i/\partial X_j\} = J$  is the Jacobian matrix. Thus, the numerical shift matrix is proportional to the negative inverse of  $J$ .

If precise numerical data cannot be obtained, qualitative assignment of the inflows causing increase (+) or decrease (−) in steady-state concentrations provide sign-symbolic  $J^{-1}$ ; this information is related to bifurcation diagrams (to be described later) and can be used to distinguish roles of essential species and to assign the category of oscillator. The sign-symbolic concentration shift matrices for the essential species in the prototype of each category are given in table 11.4. One distinct feature is immediately apparent: type Z species always has *inverse self-regulation* (negative diagonal element). Type X and Y species always have *normal self-regulation* (positive diagonal element), but can be mutually distinguished by off-diagonal elements.

Now we can apply the concentration shift regulation method to the DOP model introduced as an example in the previous section and complete the analysis. Chevalier et al. [7] show that under conditions when  $S_2 = \text{NADH}$  is nonessential,  $S_1 = \text{O}_2$  is inversely regulated, while the intermediates  $S_3$  and  $S_4$  show normal regulation. Therefore  $S_1$  is of type Z, and  $S_3, S_4$  are of type X (they are of the same type because they are mutually in phase) and the category is 2C.

Table 11.5 Effect of essential species influx on the Hopf bifurcation; an (s) indicates that added flow is stabilizing, while (d) indicates that the added flow is destabilizing

	X	Y	Z	W
1B	s	d	d	—
1CX	s	d	s	—
1CW	s	d	s	s
2B	d	—	s	—
2C	s	—	s	—

### 11.3.5 Concentration Shift Destabilization

A system near a supercritical Hopf bifurcation may be either at a steady state or oscillatory. Starting from this point, the inflow of a species is varied (increased or decreased) until the transition from oscillations to steady state or vice versa occurs. There are two possibilities:

1. Either there are oscillations on the left side of the Hopf point (and at the same time the steady state is unstable); then increase in the inflow results in the steady state and such transition is assigned as stabilizing (s).
2. Or there is a stable steady state on the left side, and increased inflow causes the steady state to become unstable and oscillations occur; this transition is called destabilizing (d).

Eiswirth et al. [1] presented a way of distinguishing roles of essential species in the mechanism and subcategories using s/d assignment in addition to self-regulation information from the concentration shift experiments (i.e., diagonal elements of  $\Delta X_{\text{symb}}$ ). Table 11.5 shows s/d patterns for each category; the crossing of the Hopf bifurcation can in general occur in two situations: either on an “upper” branch, say with high concentration of the autocatalytic species X, or on a “lower” branch. High X is assumed in the table, otherwise the s/d pattern would be just the opposite.

There is a close relation between the s/d patterns and off-diagonal elements of  $\Delta X_{\text{symb}}$ . If two columns of  $\Delta X_{\text{symb}}$  have the same signs, then the s/d pattern for the corresponding species will be either ss or dd, depending on whether the Hopf bifurcation occurs on the upper or lower branch; opposite signs in two columns of  $\Delta X_{\text{symb}}$  imply sd or ds patterns; this is so because  $\Delta X_{\text{symb}}$  is closely related to bifurcation diagrams as explained in section 11.3.6. Thus destabilization experiments may serve either as a consistency check when the entire  $\Delta X_{\text{symb}}$  is known, or as information needed to reconstruct off-diagonal elements of  $\Delta X_{\text{symb}}$ , if only diagonal elements are known.

### 11.3.6 Bifurcation Diagrams

The use of bifurcation diagrams for categorization has been initiated in Eiswirth et al. [1] and studied in detail in Schreiber and Ross [9]. There have been several attempts to use bifurcation diagrams in order to decide between proposed alternative mechanisms.

Nosztcizius et al. [53] studied the BZ reaction in a flow-through reactor with various chemical reagents being fed in and concluded that the bifurcation diagrams can be used as fingerprints of the perturbing reagent. Olsen and Epstein [54] examined the effect of alternative proposed pathways on calculated bifurcation diagrams, compared them with available experiments, and suggested further experiments for the mixed Landolt and chlorite–iodide reactions. Ringland [55] showed how certain codimension 2 points in a bifurcation diagram for a seven-variable model of the BZ reaction can be used to accept/exclude the model when compared with experimental data. For the same model, Clarke and Jiang [47] used the stoichiometric analysis to decompose the system into extreme currents. By using different current combinations they studied the effect of deleting or including particular reactions on the shape, location, and nature of the bifurcation curves. They found that the model cannot be brought into agreement with Ringland's experiments without including  $\text{Br}^-$ -consuming reactions in the network.

The task of mechanism determination may be systematically approached using the bifurcation diagrams (and closely related concentration shift matrices) constructed from experiments in a flow-through reactor, where the flow rate and inflow concentration of each species is systematically varied to provide two-parameter bifurcation diagrams. As explained in section 11.2.2, these diagrams are typically of a cross-shaped type. They contain a cusp-shaped region of multiple stationary states and an oscillatory region extending beyond the tip of the cusp; bistability is confined to the interior of the cusp and excitability often occurs near the tip. A primary feature for categorization is the tilt of the cusp-shaped region; the actual arrangement of the oscillatory region provides a supplementary feature.

Along with the bifurcation diagrams, one-parameter diagrams of steady-state concentrations near the cusp are useful in providing information about the mechanism. When the concentration of a species is plotted against a parameter across the region of bistability, then one branch can be assigned as “upper” and the other as “lower” with respect to that species. This concept can be extended to bifurcation diagrams within or near to the cusp region, including the adjacent oscillatory region. In the diagrams described below we use the label  $u$  and  $l$  to indicate relative levels of concentration. Figures 11.7 and 11.8 show the bifurcation diagrams for categories 2C, 2B and 1C, 1B respectively.

In the following, the inflow concentrations  $X_{0i}$  and the reciprocal residence time  $k_0$  are called constraints. All diagrams are plotted with the reciprocal residence time along the horizontal axis and inflow concentration of each essential species along the vertical axis. In all diagrams of this type the important feature is the orientation of the cusp with respect to the coordinate axes. Generically the cusp is tilted rather than being parallel to one of the coordinate axes. Consequently the “symmetry axis” of the cusp (approximately the line to which both branches of the cusp curve are tangent at the tip) has generically either a positive or negative slope; we call the two cases *diagonal* and *antidiagonal*, respectively. Within each category the diagonal or antidiagonal orientation is preserved for a given pair of constraints regardless of the choice of rate coefficients. Moreover, the tip of the cusp is directed specifically either to the left or to the right (with respect to the horizontal direction) for each essential species in each category. In addition, an oscillatory region may extend past the tip. In this case, the Hopf curves indicate categories and classify essential species consistently with the concentration destabilization shifts.

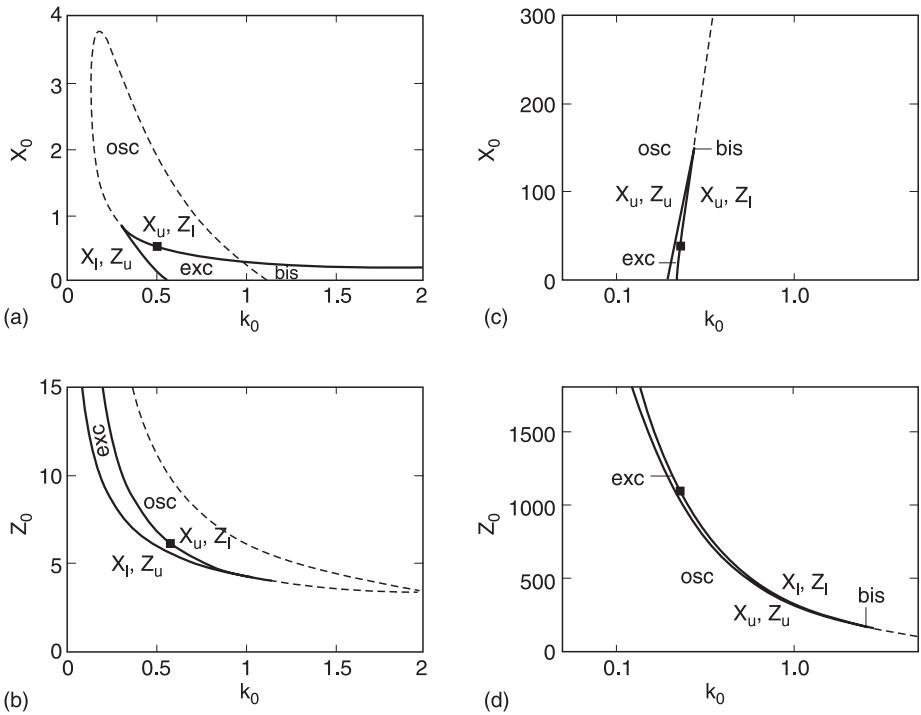


Fig. 11.7 Two-dimensional bifurcation diagrams in the parameter plane reciprocal residence time versus inflow concentration of essential species for categories 2C (a, b) and 2B (c, d). (Adapted from [9].)

Notice that a diagonal tilt implies opposite transitions for a species when followed horizontally and vertically across the region of multiplicity. For example,  $X$  in fig. 11.8(a) jumps from  $u$  to  $l$  along  $k_0$ -axis whereas the transition is from  $l$  to  $u$  along the  $X_0$ -axis. On the other hand, the antidiagonal tilt implies the same transition for each species in both axial directions, for example from  $l$  to  $u$  for  $Y$  in fig. 11.8(b). Therefore the tilt of the cusp bifurcation structure for a pair of inflow concentrations is derived from the inflow concentration/flow rate diagrams by realizing that the transition from  $u$  to  $l$  or vice versa for each species must be preserved in any bifurcation diagram. For example, if the transition for  $X$  is from  $l$  to  $u$  along the  $X_0$ -axis as in fig. 11.8(a) and from  $u$  to  $l$  along  $Y_0$ -axis as in fig. 11.8(b), then the tilt in the  $X_0$ – $Y_0$  diagram will be diagonal.

There is a straightforward relation between  $lu$  or  $ul$  transitions and concentration shifts. If we denote  $lu$  transition for the  $i$ th species with respect to change in the inflow of the  $j$ th species by  $+$  and the opposite transition by  $-$ , then we obtain a matrix equal to  $\Delta X_{\text{symp}}$ , provided that the dependence of concentration of each species on any constraint is monotonous along the upper and lower branch [9]. While the concentration shift matrix reflects local sensitivity of the steady state with respect to constraints, the bifurcation diagram is global in nature. This implies that any two columns of  $\Delta X_{\text{symp}}$  determined from the  $lu/ul$  transitions are either the same or opposite, and the corresponding tilt is

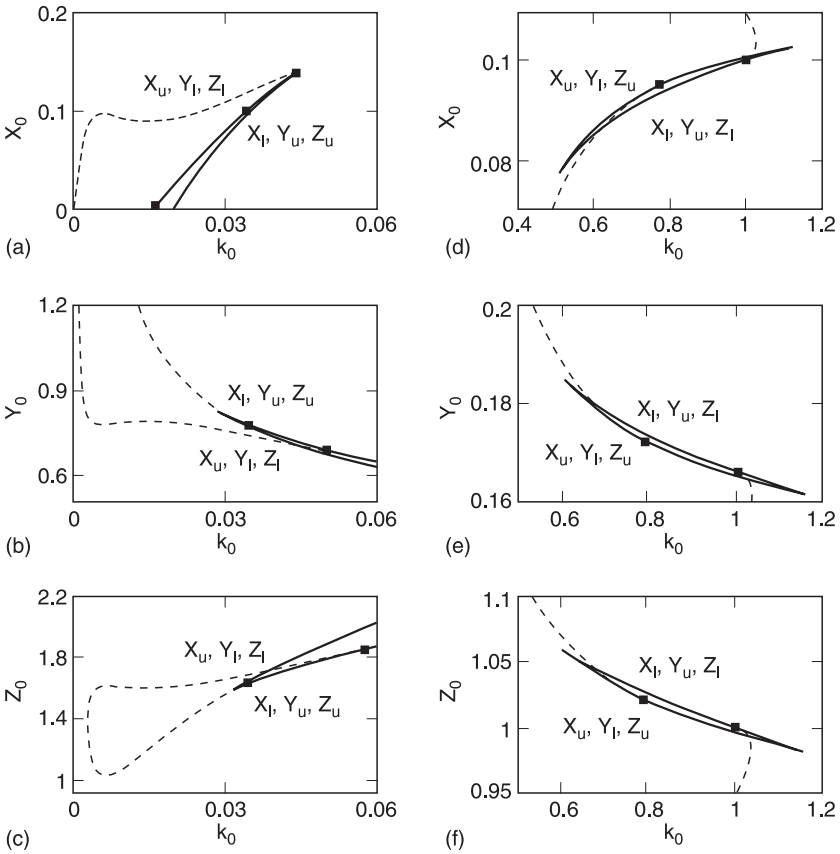


Fig. 11.8 Two-dimensional bifurcation diagrams in the parameter plane reciprocal residence time versus inflow concentration of essential species for categories 1C (a, b, c) and 1B (d, e, f). The subcategories of 1C are not distinguished and type W species is not shown since it provides the same diagram as type X species. (Adapted from [9].)

antidiagonal or diagonal, respectively. Schreiber and Ross [9] define certain relations between the rows of  $\Delta X_{\text{symp}}$  and reactions among the species involved, most noticeably, two rows with opposite signs indicate that the species react with one another (like X and Y in an exit reaction, or Z and X in category 1C). Therefore the connectivity of a mechanism can be partly resolved by the concentration shift method.

Finally, the relation between concentration destabilization shifts and the off-diagonal elements of  $\Delta X_{\text{symp}}$  is also explained via bifurcation diagrams. If the tilt is antidiagonal, then the effect of increasing either of the two constraints involved on the steady state near the Hopf bifurcation is the same (i.e., s or d, depending on whether we use the  $u$  or  $l$  steady state) for the same reasons as the two corresponding columns of  $\Delta X_{\text{symp}}$  are the same. For a diagonal tilt, the opposite holds. Thus, given the diagonal elements of  $\Delta X_{\text{symp}}$  and concentration destabilization shifts, off-diagonal elements are easily

Table 11.6 Expected responses to pulsed perturbations for essential and nonessential species of type A, B, and C

		Perturbing species			
		Essential	A	B	C
Response of measured species	Essential	Damped oscillatory	Damped oscillatory	No response	Small amplitude damped oscillatory
	A	No response	Monotonic decay	No response	No response
	B	Damped oscillatory	Damped oscillatory	Monotonic decay	Small amplitude damped oscillatory
	C	No response	No response	No response	Monotonic decay

recovered, or conversely, given the full concentration shift matrix, destabilization shifts are derived.

### 11.3.7 Pulsed Species Response

In this experiment, a pulsed perturbation of one species at a time is applied to a system at a stable steady state near a supercritical Hopf bifurcation. The relaxation to the steady state is measured. Two purposes of these experiments can be distinguished, either to discriminate among essential and nonessential species or to estimate sign-symbolic Jacobian matrix elements.

#### *Qualitative Pulsed Species Response*

The response is characterized as oscillatory, monotonic decay, or none. A nonessential species of type A shows a negligible response to perturbation by any species but itself, for which it exhibits monotonic decay; a perturbation by a type A species causes a damped oscillatory response in all essential species and in nonessential species of type B. A perturbation by a nonessential species of type B causes negligible response in all species except itself, or it shows monotonic decay. A nonessential species of type C is in time-independent quasi-steady state to which it usually returns very quickly after perturbation by any species. Finally, perturbation by essential species leads to damped oscillatory response of all essential species and of nonessential species of type B. This information is summarized in table 11.6. A complete determination of all essential and nonessential species by this method relies on measurement of most or all species and requires the addition of most or all species. However, a complete determination may be deducible from partial information, and, as seen in the example, a partial determination may be useful.

### Quantitative Pulsed Species Response

The experiments proceed as before, but the focus is on the initial phase of the response of each measurable species, and a distinction is made among three different types of responses: an increase (+), a decrease (−), or no change (0). If possible, the initial slope of the temporal evolution of each species upon perturbation is measured. The numerical Jacobian matrix (or portions of it) may be calculated according to eqs. (11.3) and (11.4) from the measured initial slopes [7]. Assuming pulsed perturbation  $\delta X_k$  by species  $k$ , the  $k$ th column of  $J$  is computed by the use the relation  $d\delta X_i/dt = (\partial F_i/\partial X_k) \delta X_k$ , where  $d\delta X_i/dt$  is the initial time slope of the measured species  $i$ . This finding can be compared to the Jacobian matrix predicted by a model reaction mechanism at the steady state.

The sign-symbolic Jacobian matrices yield information on the connectivity of the species in the reaction [2]. The diagonal elements give information about self-production and self-consumption of each species. A diagonal entry of + indicates direct autocatalysis of a species; that of — suggests no autocatalysis or, if autocatalysis of the species takes place, that the reactions consuming that species are predominant; and 0 indicates that self-production and self-consumption of the species are balanced. The off-diagonal elements give information about how the various species interact with each other.

### 11.3.8 Delay Experiments

Consider an oscillatory reaction run in a CSTR. It is sufficient that one species in the system is measured. A delayed feedback related to an earlier concentration value of one species is applied to the inflow of a species in the oscillatory system. The feedback takes the form  $X_{j0}(t) = f(X_i(t - \tau))$ , where the input concentration  $X_{j0}(t)$  of the  $j$ th species at time  $t$  depends on  $X_i(t - \tau)$ , the instantaneous concentration of the  $i$ th species at time  $t - \tau$ . Once the delayed feedback is applied, the system may remain in the oscillatory regime or may settle on a stable steady state. The delays  $\tau$  at which the system crosses a Hopf bifurcation and frequencies  $y$  of the oscillations at the bifurcation are determined. The feedback relative to a measured reference species is applied to the inflow of each of the inflow species in turn. For each species used as a delayed feedback, a number of Hopf bifurcations equal to the number of species in the system is located if possible; for a system with  $n$  species, this would provide  $n^2$  bifurcation points with different  $\tau$  and  $y$ . As described in detail by Chevalier et al. [7], experimentally determined  $\tau$  and  $y$  values can be used to find traces of submatrices of the Jacobian. If  $n$  Hopf bifurcations are located for a feedback to each of the species in the system, this allows determination of the complete Jacobian matrix from measurements of a single species in the reaction and then the connectivity of the network can be analyzed.

### 11.3.9 Quenching

The variation in time of concentrations of species in an oscillatory reaction can be stopped temporarily by application of a pulse perturbation of a species to the system. If the system is close to a supercritical Hopf bifurcation, an interpretation of the addition leading to quenching of oscillations provides information on nonessential species, and on the Jacobian matrix. In quenching experiments, the phase of oscillation at which the perturbation is added and the amount of perturbant added are varied for each perturbing

species, one at a time. The resulting quench data consist of the phase and amplitude for each species such that the oscillations are temporarily stopped. The theory underlying quench studies is detailed in Hynne et al. [48] and relies on the properties of the Jacobian matrix at the Hopf bifurcation.

As explained in section 11.2.3, emerging limit cycle oscillations are confined to a planar center eigenspace and the quenching is achieved by perturbing in the direction of a quench vector so that a stable manifold of the steady state is hit. Recall that the quench vector (along with relative amplitudes of oscillations) enables us to discriminate among essential and nonessential species (see table 11.1). Complete quench data yield a pair of complex conjugate eigenvectors of  $J^T$ , the transpose of the Jacobian matrix, associated with the pure imaginary eigenvalue of the bifurcation. This, however, is not sufficient to recover  $J$  if the number of variables is larger than two. On the other hand, quench data can be used, for example, to optimize sets of rate coefficients. Hynne and Sørensen [56] performed quench experiments on the BZ reaction. Addition of  $\text{Br}^-$ ,  $\text{Ce}^{4+}$ ,  $\text{CHBr}(\text{COOH})_2$ , and  $\text{HOBr}$  were found to quench the oscillations, but the amplitude of  $\text{CHBr}(\text{COOH})_2$  is nearly 25 times that of  $\text{Ce}^{4+}$ , and the quench amplitude of  $\text{HOBr}$  is roughly 3 times that of  $\text{Ce}^{4+}$ . Thus  $\text{CHBr}(\text{COOH})_2$  is almost certainly a nonessential species of type B or C due to the large quench amplitude. For the same reason,  $\text{HOBr}$  may also be nonessential of type B or C. The species  $\text{Br}^-$  and  $\text{Ce}^{4+}$  are identified as essential, which is consistent with the roles assigned by Eiswirth et al. [1] in a detailed model of the BZ reaction. Hynne et al. [57] also outlined a method by which to combine quenching data and representation of the BZ reaction network by extreme currents to optimize sets of rate coefficients.

### 11.3.10 Phase Response Experiments

A pulsed perturbation of one species at a time is applied to an oscillatory system, often but not necessarily near a supercritical Hopf bifurcation. The phase of oscillation at which the perturbation is added and the amount of perturbant added are varied. It is determined whether each perturbation results in an advance or delay of the next oscillatory peak relative to the period of the unperturbed autonomous system. If we let  $t_n$  denote the time of the  $n$ th reference event (an easily followed feature of the oscillations, such as a sharp rise or fall in the concentration of a monitored species), the period of the unperturbed limit cycle is given by  $T_0 = t_n - t_{n-1}$ . A pulse perturbation is applied at some time  $t_{\text{pert}}$ , between  $t_n$  and  $t_{n+1}$ , from which we define a phase of perturbation as  $\varphi_{\text{pert}} = (t_{\text{pert}} - t_{n-1})/T_0$ . Another reference event of the same type as the one at  $t_n$  follows the perturbation at time  $t'_{n+1}$  (generally different from  $t_{n+1}$  in the absence of perturbation). Perturbations for which  $t'_{n+1} > t_{n+1}$  are phase delays, while those for which  $t'_{n+1} < t_{n+1}$  are phase advances; a phase shift is then defined as  $\Delta\varphi = (t_{n+1} - t'_{n+1})/T_0$  and a new phase is  $\varphi' = \varphi_{\text{pert}} + \Delta\varphi$ . A plot of  $\Delta\varphi$  against  $\varphi_{\text{pert}}$  is a phase response curve (PRC), while a plot of  $\varphi'$  against  $\varphi_{\text{pert}}$  is a phase transition curve (PTC). In fact, there is a sequence of PRC/PTCs when successive reference events are used following the perturbation. However, usually the transient after the perturbation relaxes quickly back to the limit cycle oscillations so that the asymptotic PRC/PTC is virtually indistinguishable from that calculated from the first reference event.



The concept of the phase response in biological context has been thoroughly explored by Winfree [58]. Ruoff et al. [59] report phase response experiments in which the oscillatory BZ reaction is perturbed with  $\text{NaBrO}_2$ . They experimentally determined relations between perturbation size and the phase of perturbation at which a transition is seen between phase advances and phase delays. These observations were compared to four different models of the BZ reaction and proved useful in ruling some of them out and identifying the best one. Phase response techniques were used to study forced oscillators [60,61] and excitable systems [62]. Vance and Ross [63] point out theoretically and Millet et al. [64] show in experiments that the PRC can be used to determine the quench vector and that the phase resetting experiments may be more convenient for obtaining quench data than the quench experiments themselves. However, the use of PRC/PTC for categorization still awaits further development.

### 11.3.11 External Periodic Perturbation

A sinusoidal perturbation is applied to an inflow of species to the oscillatory system. The amplitude  $\epsilon$  and period  $T_f$  of the external forcing are varied. The entrainment boundaries of the various  $p/q$  bands (in which there are  $q$  response cycles for every  $p$  forcing cycles) are determined. The width of the entrainment bands is decreasing with  $p$  and  $q$  so it is convenient to choose the fundamental 1/1 band. As the forcing period is varied across the band, the relative phase between the perturbation and the response for each species is measured. As discussed by Eiswirth et al. [1] (see also Loud [65]), entrainment data may be used to distinguish essential from nonessential species. The total change of observed phase shift between the perturbation and the response in a particular species from one edge of the entrainment band to the other is half a cycle (i.e.,  $180^\circ$ ) for essential species, but less than that for nonessential species. In addition, nonessential species of type B require very large amplitudes for a response to be observable.

Vance and Ross [63] show that there is a close relation between the variation of the relative phase shift across the entrainment band and variation of the phase response curve provided that limit cycle oscillations are close to a supercritical Hopf bifurcation.

### 11.3.12 Stabilization of the Unstable Steady-State Branch

In the presence of multiple steady states, some branches are stable and some are unstable. Likewise, stable oscillations imply the presence of an unstable steady state from which oscillations originally emerged at a different location in the parameter space. There are experimental techniques employing a suitable (delayed) feedback stabilizing the steady state without affecting its location in the phase space. Such an investigation allows comparison of unstable branches between model and experiment. Hjelmfelt and Ross [66] report experiments in which the use of feedback to inflow rate is used to stabilize unstable steady states in the oscillatory and excitable regimes of the chlorite–iodide system. Petrov et al. [67] describe a general method for stabilizing steady states and periodic orbits in multidimensional systems. This method is applied for stabilizing nonhomogeneous steady states (Turing patterns) in reaction–diffusion systems [68].

### 11.3.13 Other Methods

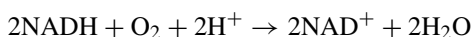
Most of the methods outlined above are suitable for obtaining information on oscillatory reaction networks. As pointed out in several other chapters in this book, related methods can be used for determination of causal connectivities of species and deduction of mechanisms in general nonoscillatory networks. Pulses of species' concentration by an arbitrary amount have been proposed (see chapter 5) and experimentally applied to glycolysis (see chapter 6). Random perturbation by a species can be used and the response evaluated by means of correlation functions (see chapter 7); this correlation metric construction method has also been tested (see chapter 8). Another approach to determining reaction mechanisms by finding Jacobian matrix elements is described in Mihaliuk et al. [69].

## 11.4 Examples of Deduction of Reaction Mechanism from Experiments

We describe here two examples of experimental systems examined with the aforementioned methods and compare the experimental results with the behavior of models based on detailed mechanisms. In each case, partial categorization is possible since only a limited number of species can be measured. Nevertheless, the results provide a deeper insight into the mechanisms and offer guidelines for further refinements of the models.

### 11.4.1 Peroxidase–Oxidase Reaction

The widely studied peroxidase–oxidase (PO) reaction exhibits a variety of different types of nonlinear dynamics. The overall reaction is



where NADH is reduced nicotinamide adenine dinucleotide. An enzyme widely used in experiments is extracted from the horseradish plant and it is commonly referred to as the horseradish peroxidase (HRP). Complex dynamics such as limit cycle oscillations [70], bistability between two steady states [71], chaos [72], quasiperiodicity [73], and mixed-mode oscillations [74] have all been observed experimentally. Apart from plants, variants of this enzyme are found in many living organisms; recently it has been found to play an important role in certain cells of the immune system [75] by virtue of its ability to produce highly oxidative radicals such as  $\text{O}_2^{\bullet-}$ , which are used to destroy agents foreign to the organism. Many different models, both abstract and detailed, have been proposed for this oscillatory reaction (see [76] for an overview).

Hung and Ross [6] carried out experiments in which some of the proposed tests are applied to assign roles to four of the detectable species in the HRP system: oxygen, NADH, the native enzyme ( $\text{Per}^{3+}$ ), and an oxidized form of the native enzyme called compound III (coIII). These tests include comparing the relative amplitude and phase shifts of the oscillations of these compounds, concentration shift regulation and destabilization experiments, qualitative pulsed species response, and quench experiments.

Table 11.7 Measurements of the relative amplitude of NADH, coIII and  $\text{Per}^{3+}$  as a percentage of the relative amplitude of oxygen

Species	% with respect to oxygen
Oxygen	100
coIII	$32 \pm 8$
$\text{Per}^{3+}$	$12 \pm 5$
NADH	$10 \pm 5$

The experimental apparatus consists of a well-stirred thermostated reactor with the reactants oxygen and NADH supplied continuously. NADH,  $\text{Per}^{3+}$ , and coIII are measured spectrophotometrically, while an oxygen electrode determines the concentration of  $\text{O}_2$ . There is no outflow from the reactor, but the change in reaction mixture volume during experiments due to inflow of NADH is negligible; pH in the solution is kept constant by using an acetate buffer and small amounts of 2,4-dichlorophenol (DCP) and methylene blue (MB) are added to stabilize the oscillations. Oxygen flows in and out of the mixture across the interface. Experimental conditions were set within a range of constraints where a supercritical Hopf bifurcation occurs and tests were done on either the stationary or the oscillatory side of the bifurcation.

### *Relative Amplitudes of Oscillations*

The relative amplitudes of the oscillating species are determined as

$$c_r = (c_i - c_a)/c_a \quad (11.12)$$

where  $c_r$  is the relative concentration,  $c_i$  is the instantaneous concentration, and  $c_a$  is the average concentration over the time period in consideration. In fig. 11.9, the relative amplitudes of the four measured species for a particular run are plotted together for comparison. Table 11.7 provides a quantitative measure of the amplitudes; oxygen is used as a gauge. The data are averaged over several experimental time series; coIII is about three-tenths, and  $\text{Per}^{3+}$  and coIII only one-tenth the value of oxygen, which suggests that the former two species are probably essential while the latter two are likely to be nonessential.

### *Phase Shifts of Oscillations*

The phase shifts are also derived from the data in fig. 11.9. Oxygen and NADH are approximately in phase during oscillations, that is, both the maxima and minima of their oscillations occur at the same time; coIII, on the other hand, has a different waveform and while its maximum is delayed relative to oxygen and NADH, its minimum occurs at about the same time. On the whole, oscillations of coIII are lagging; the native enzyme is nearly antiphase with respect to coIII, which suggests that any other enzymatic intermediates are present in small concentrations. Table 11.8 summarizes the symbolic phase shift relations.

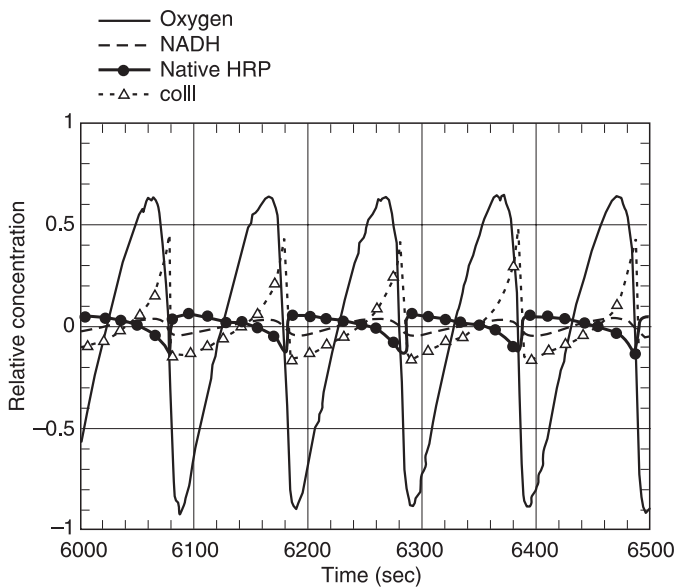


Fig. 11.9 Measurements of the relative amplitudes of  $O_2$ , NADH,  $Per^{3+}$  (native HRP), and coIII. Typical external constraints and initial conditions:  $25\ \mu\text{L/s}$  oxygen flow,  $28\ \mu\text{L/h}$  NADH flow,  $0.65\ \mu\text{M}$  of HRP,  $10\ \mu\text{M}$  of DCP,  $0.1\ \mu\text{M}$  of MB, and  $7.5\ \text{mL}$  of solution in reactor. (From [6].)

Concentration Shift Regulation

Experiments described here are performed at the middle and low stationary states of oxygen. The middle steady state refers to oxygen concentration at about half the inlet concentration, while the low steady state refers to oxygen concentrations substantially lower than the inlet concentration. Although no multiple steady states were observed in the experiments, the two levels of oxygen can be considered as corresponding to an upper and lower stationary state as discussed in section 11.3.6. Figure 11.10 shows the change in the lower stationary state with respect to oxygen due to an increase in its inflow rate.

Table 11.8 Experimental phase relations of oxygen, coIII,  $Per^{3+}$ , and NADH in the PO reaction

		Maxima of			
		Oxygen	coIII	$Per^{3+}$	NADH
With respect to	Oxygen	I	—	+	I
	coIII	+	I	A	+
	$Per^{3+}$	—	A	I	—
	NADH	I	—	+	I

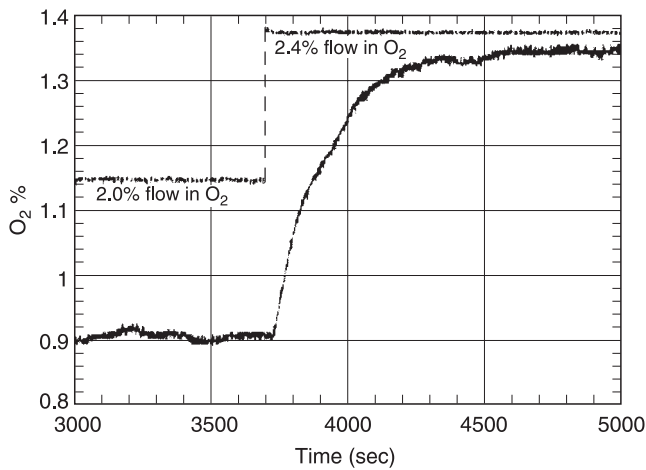


Fig. 11.10 Measurements of the concentration shift of the lower stationary state to an upper state as a result of increasing the flow rate of oxygen (dashed line) from 2.0 % to 2.4 %. External constraints and initial conditions: 2.0 mL/s total gas flow rate, 33.0  $\mu$ L/h NADH flow, 0.71  $\mu$ M of HRP, 20  $\mu$ M of DCP, 0.1  $\mu$ M of MB, and 7.5 mL of solution in the reactor. (From [6].)

The concentration shift regulation for  $\text{Per}^{3+}$  cannot be performed independently of other enzyme species because all the HRP enzyme forms are bound by a conservation constraint and therefore their regulation is mutually dependent. Table 11.9 lists the responses of the species to an increase in oxygen and NADH. The results are the same for the upper and lower stationary state.

*Concentration Shift Destabilization*

By increasing the flow of oxygen at its low stationary state across a Hopf bifurcation, the effect is destabilizing as shown in fig. 11.11. NADH is expected to be nonessential and therefore the concentration shift regulation experiments were not performed.

Table 11.9 Results of the concentration shift regulation experiments

		Increase of flow in	
		Oxygen	NADH
Measured species	Oxygen	+	–
	NADH	–	+
	$\text{Per}^{3+}$	–	+
	coIII	+	–

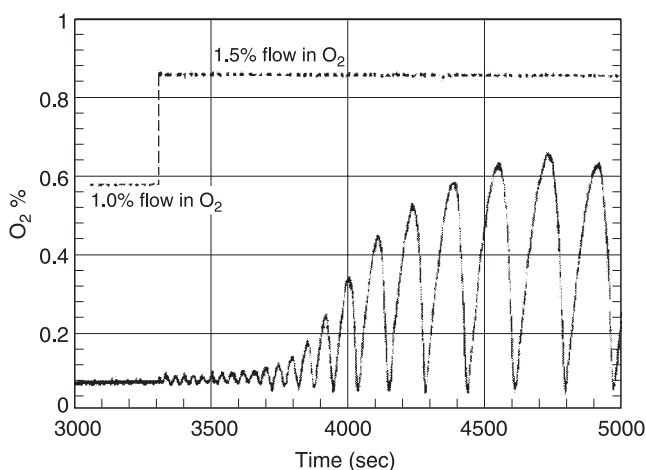


Fig. 11.11 Measurements of the transition of oxygen from its low stationary state to the oscillatory regime due to an increase in oxygen flow (dashed line) from 1.0 % to 1.5 %. External constraints and initial conditions: 2.0 mL/s total gas flow rate, 20.0  $\mu$ L/h NADH flow, 0.44  $\mu$ M of HRP, 10  $\mu$ M of DCP, 0.1  $\mu$ M of MB and 7.0 mL of solution in the reactor. (From [6].)

Nonetheless, oxygen and NADH exhibit opposite effects, since an increase in oxygen leads to an overall decrease in NADH, suggesting that the two species or their intermediates may be reacting with each other.

### *Qualitative Pulsed Species Response*

The responses of the measured species due to a pulse perturbation in one of the species are shown in table 11.10. The qualitative relaxations of the species to their stationary values provide additional information about the essential and nonessential species.

Table 11.10 Experimental responses of the relaxation of the measured species to their original stationary states due to a pulsed perturbation of oxygen, NADH, and  $\text{H}_2\text{O}_2$

		Perturbing species		
		Oxygen	NADH	$\text{H}_2\text{O}_2$
Response of measured species	Oxygen	Damped oscillations	Damped oscillations	Damped oscillations
	NADH	Small ampl. damped oscillations	Small ampl. damped oscillations	Damped oscillations
	$\text{Per}^{3+}$	Damped oscillations	Damped oscillations	Damped oscillations
	coIII	Damped oscillations	Damped oscillations	Damped oscillations

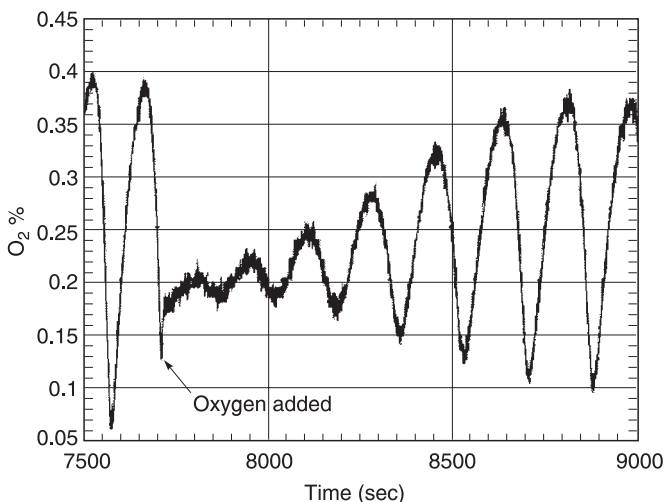


Fig. 11.12 Measurement of the quench in oscillations of the oxygen concentration due to an addition of  $1.5\ \mu\text{L}$  of air to the reacting solution at the arrow. External constraints and initial conditions:  $20.0\ \mu\text{L/s}$  oxygen flow rate,  $21.0\ \mu\text{L/h}$  NADH flow,  $0.54\ \mu\text{M}$  of HRP,  $10\ \mu\text{M}$  of DCP,  $0.1\ \mu\text{M}$  of MB and  $7.5\ \text{mL}$  of solution in the reactor. (From [6].)

### Quench Experiments

This is another method of distinguishing between essential and nonessential species of type B and C. The quench perturbation of essential species and nonessential species of type A is much smaller than that of type B and C nonessential species. Oxygen quenches the oscillations temporarily if a small amount of it is added at a phase when its concentration is decreasing (see fig. 11.12). Adding varying amounts of NADH to the system at different oscillatory phases does not quench the oscillations and confirms their assignments.

### Discussion

In beginning the assignment of the roles of species and the categorization of the oscillator, we can immediately eliminate certain categories: 1B and 2B may be excluded from consideration since the reaction does not exhibit sustained oscillatory behavior under batch conditions, only damped oscillations. A summary of the conclusions as to the assignment of the types of the measured species is given in table 11.11. The experiments indicate that oxygen is an essential species according to its relative amplitude and the pulse perturbation and quench experiments. This provides more evidence that the category is not B since one of its inflow species is essential. CoIII has the next largest relative amplitude, and combining this result with the pulsed perturbation experiments, we conclude that it is also an essential species. Assuming a category C, the phase shift relations and the concentration shift experiments suggest that oxygen is an essential species of type Y and coIII is of type Z, since the oscillatory maximum of oxygen occurs earlier than that of coIII. This agrees only with the suggested assignment because the other option,

**Table 11.11** Assignment of roles of species and their types by means of the results of the different categorization tests

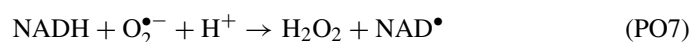
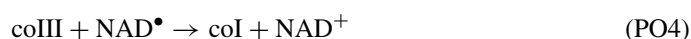
Test	Species			
	Oxygen	NADH	Per <sup>3+</sup>	coIII
Relative amplitude	E	NE	NE	E
Relative phases	E of type Y	NA	NA	E of type Z
Concentration shift regulation	E of type Y	NA	NA	E of type Z
Qualitative pulsed species response	E	E or NE	NA	E
Quenching	E or NE of type A	NE	NA	NA
Conclusion	E of type Y	NE	NE	E of type Z

E, essential species; NE, nonessential species; NA, not applicable.

that oxygen is of type Z advancing coIII of type X, is ruled out by measured normal regulation of oxygen. At the same time, this assignment implies category 1C rather than 2C because type Y species is present. NADH and Per<sup>3+</sup> are nonessential owing to their small relative amplitudes and the inability of NADH to quench the oscillations. They are both likely to be of type C, because of their supposed roles as intermediates and a large quench amplitude of NADH. It remains to be determined whether the category is 1CX or 1CW. The difference is subtle: inflow of both X and Y is required for 1CX, whereas only inflow of Y is sufficient in 1CW. Since the only essential species that is flowed in is oxygen, we conclude that the peroxidase–oxidase system is a 1CW oscillator.

### Model

The simplest of the realistic mechanisms for the PO system is model A by Aguda and Clarke [77]. The reaction steps in model A are as follows:





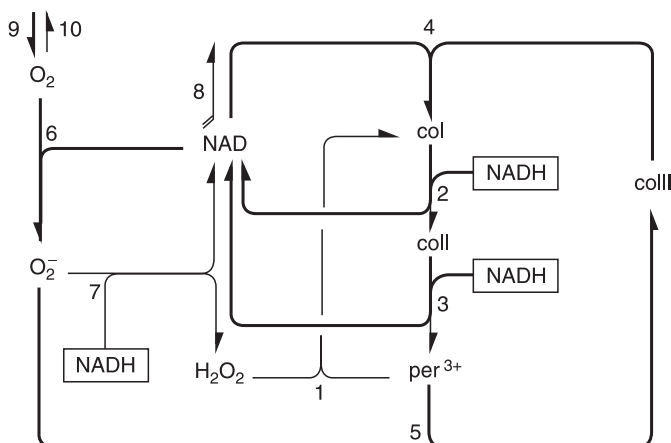


Fig. 11.13 Network diagram for model A of the PO reaction. The unstable subnetwork is shown in thick lines. (From [8].)



NADH is assumed effectively constant,  $\text{H}^+$  is buffered,  $\text{NAD}^+$  and  $(\text{NAD})_2$  are products, and therefore all these species are nonessential. As opposed to experiments, NADH is assumed a priori nonessential in the model. The variable species include the four enzyme forms,  $\text{Per}^{3+}$ ,  $\text{coI}$ ,  $\text{coII}$ , and  $\text{coIII}$ , two free radical species,  $\text{NAD}^\bullet$  and  $\text{O}_2^{\bullet-}$ , and oxygen and hydrogen peroxide. Since the enzyme is not flowed in and out, there is a conservation constraint for all the enzyme species.

Schreiber et al. [8] employed the stoichiometric network analysis outlined in section 11.2.4 to analyze sources of oscillatory instabilities in this model (and some other models of the PO system). The reaction network diagram for model A is shown in fig. 11.13. There are three extreme subnetworks (elementary modes) involving reactions: (1) PO9, PO10, (2) PO2–PO6, PO9, (3) PO1–PO3, PO6–PO9. Assuming mass action kinetics, the stability of stationary states in these subnetworks is indicated by calculating the corresponding principal subdeterminants  $\beta_l$  of the matrix  $B$  in eq. (11.10). While the stationary states of the first and third subnetworks are always stable, the second subnetwork has a negative  $\beta_4$  associated with the species  $\text{coI}$ ,  $\text{coII}$ ,  $\text{NAD}^\bullet$ , and  $\text{O}_2$  and two  $\beta_5$ 's involving the previous four species plus  $\text{Per}^{3+}$  and  $\text{O}_2^{\bullet-}$ . According to the SNA, this indicates that the subnetwork (2) may have unstable stationary states provided that the concentrations of the species associated either with  $\beta_4$  or with some of the two  $\beta_5$ 's are small enough. The reactions of the subnetwork (2) are shown in thick lines. Clearly, there is a positive feedback loop connecting cyclic species  $\text{coI}$ ,  $\text{coII}$ , and  $\text{NAD}^\bullet$  (type X species) combined with an exit reaction PO6 that involves  $\text{O}_2$  (type Y species). Those are the species involved in  $\beta_4$ . Furthermore, considering one of the negative  $\beta_5$ 's,  $\text{Per}^{3+}$  and  $\text{O}_2^{\bullet-}$  would be of type W if they prove to be essential because they are neither cyclic nor exit species. A type Z species needed for oscillations

is not indicated by a negative  $\beta_l$  since it has to be provided in relatively large amounts. By inspection of the network it is easy to see that the only suitable choice for the type Z species is coIII since it controls the autocatalytic cycle via reaction PO4.

This general analysis provides useful insights but needs to be verified for a given set of rate coefficients and constraints by calculating the quantities corresponding to (some of) the experimental tests for classification. For example, we can vary constraints (inflows of the nonenzymatic species in this case) so as to find a supercritical Hopf bifurcation, then calculate the Jacobian matrix and determine (1) its inverse to find concentration shifts, (2) the right eigenvector associated with the bifurcating pure imaginary eigenvalue to determine relative amplitudes and phase shifts, and (3) the left eigenvector to calculate quench vectors.

Calculations of relative amplitudes, quench vectors, and buffering of the species indicate that  $\text{Per}^{3+}$  and  $\text{H}_2\text{O}_2$  are nonessential species. Since the enzymatic species are bound by a conservation constraint (their overall concentration is constant), the Jacobian matrix is singular and cannot be inverted. However, this can be circumvented by calculating the inverse of an auxiliary matrix [8], which provides concentration shifts with respect to only those species that are not involved in the conservation constraint. Leaving out the nonessential species, the concentration shifts with respect to admissible species ( $\text{NAD}^\bullet$ ,  $\text{O}_2^{\bullet-}$ , and  $\text{O}_2$ ) along with phase shifts relative to  $\text{NAD}^\bullet$  (type X species) are summarized in table 11.12.

As expected for cycle species, the phase shifts of coI and coII with respect to  $\text{NAD}^\bullet$  are small, confirming that they are of type X. CoIII is advancing  $\text{NAD}^\bullet$  by about one-third of the cycle and therefore is consistent with type Z. Oxygen is close to antiphase relative to  $\text{NAD}^\bullet$  and thus is of type Y as suggested by the analysis of the network. Moreover, the phase shift of  $\text{O}_2^{\bullet-}$  is small enough to indicate that it is indeed of type W, because it is not a part of the autocatalytic cycle yet behaves like a type X species. Sign-symbolic concentration shifts with respect to  $\text{NAD}^\bullet$  and  $\text{O}_2^{\bullet-}$  are the same, which confirms the assignment of  $\text{O}_2^{\bullet-}$  as a type W species. The shifts with respect to  $\text{O}_2$  are inconclusive because the sign-symbolic representation should ideally be either the same as or opposite

Table 11.12 Phase shifts and sign-symbolic concentration shifts for the model A at an arbitrarily selected Hopf bifurcation point

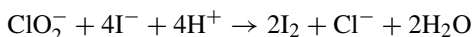
		Phase shift with respect to $\text{NAD}^\bullet$	Concentration shift with respect to		
			$\text{NAD}^\bullet$	$\text{O}_2^{\bullet-}$	$\text{O}_2$
Measured species	$\text{NAD}^\bullet$	$0^\circ$	+	+	+
	$\text{O}_2^{\bullet-}$	$19^\circ$	+	+	+
	$\text{O}_2$	$162^\circ$	—	—	+
	coI	$9^\circ$	+	+	+
	coII	$8^\circ$	+	+	+
	coIII	$128^\circ$	—	—	+

to other columns of the matrix. This is a consequence of the local determination of the Jacobian at a particular Hopf bifurcation point. A global method, such as construction of bifurcation diagrams, is needed to resolve this ambiguity. Nevertheless, the overall evidence from the analysis of the model A leads to the conclusion that it is a 1CW oscillator with coI, coII, and NAD<sup>•</sup> as type X species, O<sub>2</sub><sup>•-</sup> as type W species, O<sub>2</sub> as type Y species, and coIII as type Z species (note that coIII is produced internally rather than being supplied in the feed as in the prototype of 1CW). The remaining oscillating species Per<sup>3+</sup> and H<sub>2</sub>O<sub>2</sub> are nonessential of type C (because they are intermediates).

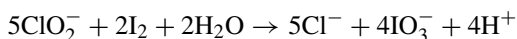
When these results are compared to the experimental findings, agreement is found in the behavior of oxygen as well as Per<sup>3+</sup> and coIII. The nonessential nature of NADH is implicitly assumed in the model and is corroborated by the experiments. Therefore the model seems to capture the essential features, even though more elaborate models need to be formulated to account for all the subtle phenomena, such as exact waveforms, and the ability to reproduce chaotic and other complex dynamics [76,78].

#### 11.4.2 Chlorite–Iodide Reaction

The chlorite-iodide reaction has been found by Dateo et al. [79] and later by De Kepper et al. [80] to exhibit complex phenomena such as oscillations and bistability between steady states, when run in a continuous-flow stirred tank reactor (CSTR). The overall stoichiometry of the reaction is



Chlorite participates in a further reaction with iodine which has the overall stoichiometry



Stemwedel and Ross [3,4] applied experimentally some categorization tests in order to assign the roles of the three detectable species in the system (I<sup>-</sup>, ClO<sub>2</sub><sup>-</sup>, and I<sub>2</sub>) and of the additional species, which can be used to perturb the system (NaOCl, IO<sub>3</sub><sup>-</sup>, Cl<sup>-</sup>, HIO<sub>2</sub>, and H<sub>2</sub>OI<sup>+</sup>). The tests applied include qualitative pulsed species response experiments, concentration shift regulation and destabilization, comparison of phase shifts of the detectable species, and bifurcation diagrams. The chlorite–iodide reaction was run in a thermostated CSTR with the inflow of solutions of chlorite, iodide, and a buffering solution. Concentration of iodide was measured with an iodide ion-selective electrode, while concentration of chlorite and iodine was determined spectrophotometrically.

#### *Qualitative Pulsed Species Response*

The iodide potential in response to the addition of pulse perturbations in I<sup>-</sup>, ClO<sub>2</sub><sup>-</sup>, Cl<sup>-</sup>, I<sub>2</sub>, HIO<sub>2</sub>, and H<sub>2</sub>OI<sup>+</sup> (protonated form of HOI) was measured. The observed responses of the I<sup>-</sup> evolution are shown in table 11.13. Since a pulsed perturbation in I<sup>-</sup> leads to a damped oscillatory response in itself, I<sup>-</sup> is an essential species. The species whose perturbations produce no response in I<sup>-</sup>, that is, IO<sub>3</sub><sup>-</sup> and I<sub>2</sub>, are then nonessential species of type B, and the one producing small amplitude damped oscillatory decay, Cl<sup>-</sup>, is most likely a nonessential species of type C. The species that are monitored but whose perturbations cause damped oscillations in I<sup>-</sup>, that is, ClO<sub>2</sub><sup>-</sup>, HIO<sub>2</sub>, and H<sub>2</sub>OI<sup>+</sup>, are

**Table 11.13** Results of qualitative species response experiments monitoring iodide potential at a stable stationary state near a Hopf bifurcation

Perturbing species	Response
$I^-$	Damped oscillations
$ClO_2^-$	Damped oscillations
$Cl^-$	Small amplitude damped oscillations
$IO_3^-$	No response
$I_2$	No response
$H_2OI^+$	Damped oscillations
$HIO_2$	Damped oscillations

either essential species or nonessential species of type A (more likely essential because of their assumed active involvement in the mechanism).

### *Concentration Shift Regulation*

Experiments were performed at an intermediate iodide stationary state near the supercritical Hopf bifurcation. The results are summarized in table 11.14. We find that  $I^-$  displays normal self-regulation, while  $ClO_2^-$  displays inverse self-regulation. Thus,  $ClO_2^-$  is the essential species of type Z. Since oscillations in this reaction are flow-dependent, we can assume that the categories 1B and 2B are ruled out. We can also rule out category 2C based on off-diagonal elements of the sign-symbolic shift matrix. In particular, the inflow of  $I^-$  causes an increase in the steady-state concentration of  $ClO_2^-$ , a feature that is not found in the corresponding prototype matrix for category 2C but occurs in that for category 1CX or 1CW. At the same time, this signifies that  $I^-$  is of type Y rather than X.

### *Concentration Shift Destabilization*

The experiments were performed near the supercritical Hopf bifurcation where the system is on the “upper” branch with respect to iodide; the system was monitored with an iodide-selective electrode. Upon addition of constant perturbing inflows of  $ClO_2^-$ , NaOCl, and  $I^-$ , at the oscillatory side of the Hopf bifurcation, the system relaxes to a

**Table 11.14** Results of concentration shift experiments obtained by monitoring iodide potential and chlorite absorbance

		Increase of flow in		
		$ClO_2^-$	$I^-$	$ClO^-$
Measured species	$ClO_2^-$	—	+	—
	$I^-$	—	+	—

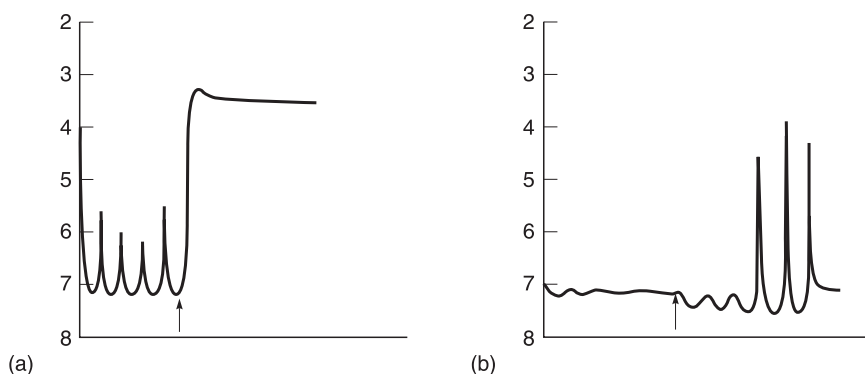


Fig. 11.14 Concentration shift destabilization experiments. Traces of iodide potential versus time are shown. (a) System starts on the oscillatory side of a supercritical Hopf bifurcation. When an inflow of NaOCl is added, the system shifts to the low iodide steady state. (b) System starts on the steady-state side of the Hopf bifurcation. When an additional inflow of  $\text{I}^-$  is applied, the system shifts to the oscillatory regime. (From [2].)

stable steady state (s). For example, for a small increase in concentration of NaOCl the system ceases to oscillate, for larger perturbations the system jumps onto the “lower” branch, in other words, an additional jump across a saddle-node bifurcation occurs, as shown in fig. 11.14(a) (this  $ul$  transition is equivalent to the stabilizing effect; see section 11.3.6). Addition of a constant perturbing inflow of  $\text{I}^-$  to the steady state causes the dynamics to become oscillatory (d) (see fig. 11.14b).

The results of all such experiments are summarized in table 11.15. As discussed above, the chlorite–iodide system is not oscillating under batch conditions and categories 1B and 2B can be safely ruled out. By inspecting s/d patterns for categories 1C and 2C in table 11.5,  $\text{I}^-$  cannot be of type X or W because it is destabilizing. In conclusion, the iodide ion must be of type Y and the category is 1C. We already know that  $\text{ClO}_2^-$  is of type Z (inverse concentration self-regulation), which cannot be determined from the s/d patterns alone. In addition, NaOCl is stabilizing, which suggests that HOCl may be type X or W species, provided that it is essential (this is important for comparison with a model discussed below).

Table 11.15 Results of concentration shift destabilization experiments obtained by monitoring iodide potential

Perturbing species	Assignment
$\text{I}^-$	Destabilizing
$\text{ClO}_2^-$	Stabilizing
NaOCl	Stabilizing
$\text{I}_2$	Stabilizing

### Bifurcation Diagrams

Figure 11.15 is an experimentally generated bifurcation diagram with reciprocal residence time  $k_0$  on the abscissa and the input concentration ratio  $[\text{ClO}_2^-]_0/[\text{I}^-]_0$  on the ordinate (this figure should be compared against a corresponding calculated bifurcation diagram in fig. 11.1). The region of oscillations is bounded from the left up by a supercritical Hopf bifurcation, and from the right down by a saddle-node infinite period bifurcation, which is simultaneously an upper branch of the cusp region. Excitability found below this branch falls into the cusp region of multiple steady states near its tip. From experiments by Luo and Epstein [81] it follows that the bistability region and hysteresis occur for higher values of both  $k_0$  and  $[\text{ClO}_2^-]_0/[\text{I}^-]_0$ . Therefore the tip of the cusp region is expected to point in the opposite (i.e., left-down) direction.

The tilt of this bifurcation structure is diagonal and can be used to explain two effects. First, if  $[\text{I}^-]_0$  is held fixed and  $[\text{ClO}_2^-]_0$  is increased, this diagram is consistent only with that for type Z species in category 1C (diagonal tilt with the tip of the cusp region extending in the left-down direction). When  $[\text{ClO}_2^-]_0$  is held fixed and  $[\text{I}^-]_0$  is increased, the tilt becomes antidiagonal, which is consistent with type Y species. Therefore the corresponding columns in the sign-symbolic concentration shift matrix are of opposite signs (see table 11.14) and the bifurcation diagram in the  $[\text{ClO}_2^-]_0$ – $[\text{I}^-]_0$  plane should be diagonal (as indeed shown in De Kepper et al. [80]). Moreover, the Hopf bifurcation is crossed from oscillatory to stable steady state when  $[\text{ClO}_2^-]_0$  is increased, that is, this species is stabilizing, while the opposite is true for  $[\text{I}^-]_0$ . The opposite s/d pattern for

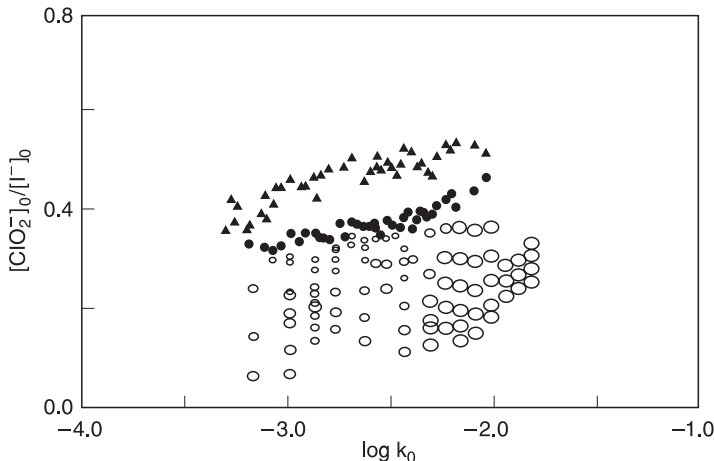


Fig. 11.15 The experimentally determined bifurcation structure of the chlorite-iodide reaction. The two constraints used here are the ratio of input concentrations,  $[\text{ClO}_2^-]_0/[\text{I}^-]_0$ , and the logarithm of the reciprocal residence time,  $\log k_0$ . Notation: filled triangles, supercritical Hopf bifurcations; filled circles, saddle-node infinite period bifurcations; in the region between these two kind of bifurcations, stable oscillations were observed. Open circles, excitable dynamics; the smallest circles correspond to perturbations of  $2 \times 10^{-3}$  M in  $\text{NaClO}_2$ ; the next smallest,  $6 \times 10^{-3}$  M in  $\text{NaClO}_2$ ; the next smallest,  $1.25 \times 10^{-3}$  M in  $\text{AgNO}_3$ ; and the largest,  $2 \times 10^{-2}$  M in  $\text{AgNO}_3$ . (From [3].)

the two species deduced from the bifurcation diagram is also corresponding to columns for  $I^-$  and  $ClO_2^-$  of opposite sign in the sign-symbolic concentration shift matrix in table 11.14. This is an example of relations among bifurcation diagrams, concentration shift regulation, and concentration shift destabilization described in section 11.3.6.

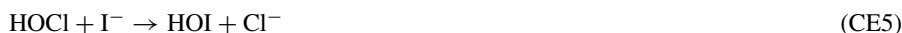
### *Phase Shifts of Oscillations*

For the chlorite–iodide system the maximum amplitude of the  $[ClO_2^-]$  and  $[I_2]$  are found to be in phase, while  $[I^-]$  is advancing both species by a phase smaller than  $180^\circ$ . Knowing that  $ClO_2^-$  is an essential species of type Z,  $I^-$  must be an essential species of type Y in category 1C, a result consistent with all the previous tests.

In conclusion, from all the experiments we deduce consistently that the chlorite–iodide system is a category 1C oscillator, where  $ClO_2^-$  is an essential species of type Z and  $I^-$  is an essential species of type Y. Further,  $HOCl$  is likely an essential species of type X or W, and  $HOI$  and  $HIO_2$  are probably essential with an undetermined role.  $I_2$  and  $IO_3^-$  are nonessential species of type B, while  $Cl^-$  is nonessential of type C.

### *Model*

Epstein and Kustin [39] set up a first reaction mechanism for the chlorite–iodide oscillator and reported a (diagonal) cross-shaped diagram in the species  $[ClO_2^-]_0$ – $[I^-]_0$ . Later, Citri and Epstein [40] revised the first suggestion for the mechanism by reducing the original model considerably. Here we employ the revised mechanism proposed by Citri and Epstein. The elementary steps are as follows:



where the dependence of the rate laws of reactions CE1, CE2, and CE6 on the buffered concentration of  $H^+$  is included in the rate coefficients.  $HClO_2$  and  $ClO_2^-$  are considered to be the same variable because of a rapid equilibration. In addition to chemical reactions CE1–CE10, pseudoreactions representing the inflow of  $I^-$  and  $ClO_2^-$  and the outflow of all species are taken into account. Evolution equations in the form of eqs. (11.8) for the dynamical variables  $HClO_2$ ,  $ClO_2^-$ ,  $HOCl$ ,  $HIO_2$ ,  $HOI$ ,  $I^-$ , and  $I_2$  are based on mass balances and mass action kinetics.

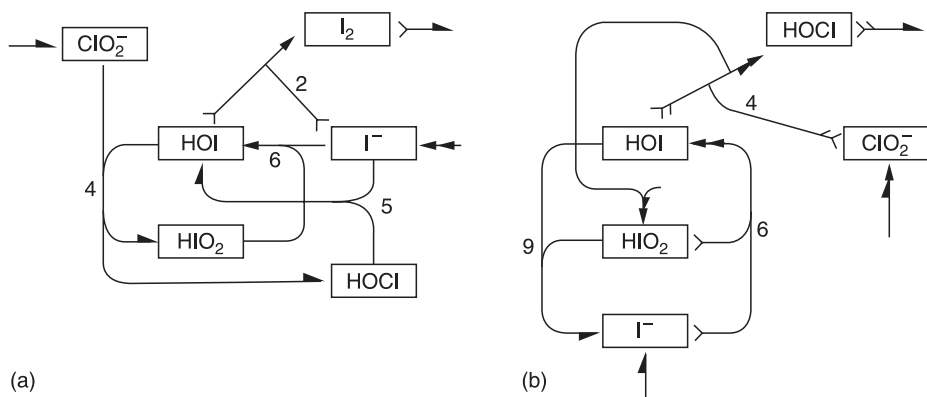


Fig. 11.16 Network diagrams of the most important extreme subnetworks of the Citri–Epstein mechanism. Subnetwork (a) possesses an unstable feature with  $\text{HOI}$  and  $\text{HIO}_2$  as cycle species and  $\text{I}^-$  as the exit species, while in the subnetwork (b)  $\text{HOI}$  and  $\text{I}^-$  are the cyclic species and  $\text{ClO}_2^-$  is the exit species. (From [5].)

Strasser et al. [5] used the SNA to identify unstable subnetworks leading to oscillatory dynamics and determine possible role of species involved. Out of the total of 39 extreme subnetworks there are 7 unstable ones. These unstable subnetworks contain essentially two basic features generating an instability, and two representative subnetworks possessing these two different features are shown in fig. 11.16. On combining either type of unstable subnetwork with appropriate stable subnetworks, a Hopf bifurcation can be obtained. By calculating the principal subdeterminants  $\beta_l$  as in the previous example, we find that the instability of the first type (fig. 11.16a) is mainly based on the interplay of  $\text{I}^-$ ,  $\text{HOI}$ , and  $\text{HIO}_2$ . Here,  $\text{HOI}$  and  $\text{HIO}_2$  form an autocatalytic cyclic pathway (type X species) via CE4 and CE6, while  $\text{I}^-$  is involved in an exit reaction CE2.  $\text{HOCl}$  is also indicated as essential, and since it is a byproduct of the autocatalysis, its suggested role is that of a type W species. Finally, with the use of the Jacobian matrix at the Hopf bifurcation, we find that  $\text{ClO}_2^-$  is controlling the cycle via a negative feedback, and therefore plays the role of type Z species. On the other hand, in the second type (fig. 11.16b) the autocatalytic species are  $\text{HOI}$  and  $\text{I}^-$ , while  $\text{ClO}_2^-$  is the exit species and  $\text{HIO}_2$  is the type Z species.

This analysis was done without assuming any values of rate coefficients and represents the general possibilities of the network. With given rate coefficients according to Citri and Epstein [40], only one of the options dominates because compound oscillations contributed from different oscillatory sources were observed neither in calculations nor in measurements. For an arbitrarily chosen stationary state near a Hopf bifurcation, Strasser et al. [5] calculated a relative percent contribution of the most pertinent extreme subnetworks. The result is that only the first type of instability is involved in the model. Since the assignment of species associated with this instability is in complete agreement with experiment, we conclude that the model with the given rate coefficients is consistent with experiments and we identify the category as 1CW with  $\text{HOCl}$  as the type W species.



## 11.5 Limits of Stoichiometric Network Analysis

As this example shows, stoichiometric network analysis yields only possibilities of reaction mechanisms, it does not lead to the determination of which of many possible reaction pathways corresponds to experiments. We need to know the rate coefficients of the system to determine paths in the system by which reactants proceed to form products. This distinction has been overlooked in a number of studies on complex biochemical reaction mechanisms (see Schilling et al. [82] and references therein).

**Acknowledgments** This chapter is based mainly on work published in refs. [1–9], where the ideas of the stoichiometric network analysis [10] were utilized to identify several distinct topological features in chemical networks that provide oscillatory instabilities and from that derive a classification system for chemical oscillators and species taking part in these oscillations.

## References

- [1] Eiswirth, M.; Freund, A.; Ross, J. Mechanistic classification of chemical oscillators and the role of species. *Adv. Chem. Phys.* **1991**, *90*, 127–199; Operational procedure toward the classification of chemical oscillators. *J. Phys. Chem.* **1991**, *95*, 1294–1299.
- [2] Stemwedel, J. D.; Schreiber, I.; Ross, J. Formulation of oscillatory reaction-mechanisms by deduction from experiments. *Adv. Chem. Phys.* **1995**, *89*, 327–388.
- [3] Stemwedel, J.; Ross, J. Experimental determination of bifurcation features of the chlorite iodide reaction. *J. Phys. Chem.* **1993**, *97*, 2863–2867.
- [4] Stemwedel, J.; Ross, J. New measurements on the chlorite iodide reaction and deduction of roles of species and categorization. *J. Phys. Chem.* **1995**, *99*, 1988–1994.
- [5] Strasser, P.; Stemwedel, J. D.; Ross, J. Analysis of a mechanism of the chlorite iodide reaction. *J. Phys. Chem.* **1993**, *97*, 2851–2862.
- [6] Hung, Y.-F.; Ross, J. New experimental methods toward the deduction of the mechanism of the oscillatory peroxidase oxidase reaction. *J. Phys. Chem.* **1995**, *99*, 1974–1979.
- [7] Chevalier, T.; Schreiber, I.; Ross, J. Toward a systematic determination of complex-reaction mechanisms. *J. Phys. Chem.* **1993**, *97*, 6776–6787.
- [8] Schreiber, I.; Hung, Y.-F.; Ross, J. Categorization of some oscillatory enzymatic reactions. *J. Phys. Chem.* **1996**, *100*, 8556–8566.
- [9] Schreiber, I.; Ross, J. Mechanisms of oscillatory reactions deduced from bifurcation diagrams. *J. Phys. Chem.* **2003**, *107*, 9846–9859.
- [10] Clarke, B. L. Stability of complex reaction networks. *Adv. Chem. Phys.* **1980**, *43*, 1–278.
- [11] Gray, P.; Griffiths, J. F.; Hasko, S. M.; Lignola, P. G. Oscillatory ignitions and cool flames accompanying the non-isothermal oxidation of acetaldehyde in a well stirred, flow reactor. *Proc. Roy. Soc. Lond. A* **1981**, *374*, 313–339.
- [12] Skrumeda, L. L.; Ross, J. Further measurements on the oscillatory cool flame oxidation of acetaldehyde and comparison with reaction-mechanism models. *J. Phys. Chem.* **1995**, *99*, 12835–12845.
- [13] Chang, M.; Schmitz, R. A. An experimental study of oscillatory states in a stirred reactor. *Chem. Eng. Sci.* **1975**, *30*, 21–34.
- [14] Puszynski, J.; Hlavacek, V. Experimental study of ignition and extinction waves and oscillatory behavior of a tubular nonadiabatic fixed-bed reactor for the oxidation of carbon monoxide. *Chem. Eng. Sci.* **1984**, *39*, 681–692.
- [15] Belousov, B. P. A periodic reaction and its mechanism. In *Sbornik referatov po radiatsionnoj medicene za 1958 g.* (conference proceedings); Medgiz: Moscow, 1959, p 145.
- [16] Zhabotinsky, A. M. Periodical oxidation of malonic acid in solution (a study of the Belousov reaction kinetics). *Biofizika* **1964**, *9*, 306–311.

- [17] Field, R. J.; Körös, E.; Noyes, R. M. Oscillations in chemical systems. II. Thorough analysis of temporal oscillation in the bromate-cerium-malonic acid system. *J. Am. Chem. Soc.* **1972**, *94*, 8649–8664.
- [18] Feinberg, M. Complex balancing in general kinetic systems. *Arch. Rat. Mech. Anal.* **1972**, *49*, 187–194; Chemical-reaction network structure and the stability of complex isothermal reactors. 1. The deficiency-zero and deficiency-one theorems. *Chem. Eng. Sci.* **1987**, *42*, 2229–2268; Chemical-reaction network structure and the stability of complex isothermal reactors. 2. Multiple steady-states for networks of deficiency one. *Chem. Eng. Sci.* **1988**, *43*, 1–25.
- [19] Luo, Y.; Epstein, I. R. Systematic design of chemical oscillators. 58. Feedback analysis of mechanisms for chemical oscillators. *Adv. Chem. Phys.* **1990**, *79*, 269–299.
- [20] Epstein, I. R.; Pojman, J. A. *Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns and Chaos*; Oxford University Press: New York, 1998.
- [21] Gray, P.; Scott, S. K. *Chemical Oscillations and Instabilities: Nonlinear Chemical Kinetics*; Oxford University Press: Oxford, 1990.
- [22] Chance, B.; Pye, E. K.; Ghosh, A. K.; Hess, B., Eds. *Biological and Biochemical Oscillator*; Academic Press: New York, 1973.
- [23] Goldbeter, A.; Caplan, S. R. Oscillatory enzymes. *Annu. Rev. Biophys. Bioeng.* **1976**, *5*, 449–476.
- [24] Rapp, P. E. An atlas of cellular oscillators. *J. Exp. Biol.* **1979**, *81*, 281–306.
- [25] Rapp, P. E.; Berridge, M. J. The control of trans-epithelial potential oscillations in the salivary gland of *Calliphora erythrocephala*. *J. Exp. Biol.* **1981**, *93*, 119–132.
- [26] Ross, J.; Schell, M. Thermodynamic efficiency in nonlinear biochemical reactions. *Annu. Rev. Biophys. Biophys. Chem.* **1987**, *16*, 401–422.
- [27] Goldbeter, A. *Biochemical Oscillations and Cellular Rhythms*; Cambridge University Press: Cambridge, 1996.
- [28] Aidley, D. J. *The Physiology of Excitable Cells*; Cambridge University Press: Cambridge, 1998.
- [29] Lauffenburger, D. A.; Lindermann, J. J. *Receptors: Models for Binding, Trafficking, and Signalling*; Oxford University Press: New York, 1993.
- [30] Sel'kov, E. E. Stabilization of energy charge, generation of oscillation and multiple steady states in energy metabolism as a result of purely stoichiometric regulation. *Eur. J. Biochem.* **1975**, *59*, 151–157.
- [31] Rapp, P. E.; Mees, A. I.; Sparrow, C. T. Frequency encoded biochemical regulation is more accurate than amplitude dependent control. *J. Theor. Biol.* **1981**, *90*, 531–544.
- [32] Keener, J.; Sneyd, J. *Mathematical Physiology*; Springer-Verlag: New York, 1998.
- [33] Fall, C. P.; Marland, E. S.; Wagner, J. M.; Tyson, J. J., Eds. *Computational Cell Biology*; Springer-Verlag: New York, 2002.
- [34] Meyer, T.; Stryer, L. Calcium spiking. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 153–174.
- [35] Dupont, G.; Goldbeter, A. One-pool model for  $\text{Ca}^{2+}$  oscillations involving  $\text{Ca}^{2+}$  and inositol 1,4,5-trisphosphate as co-agonists for  $\text{Ca}^{2+}$  release. *Cell Calcium* **1993**, *14*, 311–322.
- [36] Smolen, P.; Baxter, D. A.; Byrne, J. H. Modeling transcriptional control in gene networks—methods, recent results, and future directions. *Bull. Math. Biol.* **2000**, *62*, 247–292.
- [37] Hasty, J.; Isaacs, F.; Dolnik, M.; McMillen, D.; Collins, J. J. Designer gene networks: towards fundamental cellular control. *Chaos* **2001**, *11*, 207–220.
- [38] Tyson, J. J. Classification of instabilities in chemical reaction systems. *J. Chem. Phys.* **1975**, *62*, 1010–1015.
- [39] Epstein, I. R.; Kustin, K. Systematic design of chemical oscillators. 27. A mechanism for dynamical behavior in the oscillatory chlorite iodide reaction. *J. Phys. Chem.* **1985**, *89*, 2275–2282.
- [40] Citri, O.; Epstein, I. R. Systematic design of chemical oscillators. 42. Dynamic behavior in the chlorite iodide reaction—a simplified mechanism. *J. Phys. Chem.* **1987**, *91*, 6034–6040.
- [41] Guckenheimer, J.; Holmes, P. *Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields*; Springer-Verlag: New York, 1983.

- [42] Doedel, E. J.; Keller, H. B.; Kernévez, J. P. Numerical analysis and control of bifurcation problems: (I) Bifurcation in finite dimensions. *Int. J. Bifurcation & Chaos* **1991**, *1*, 493–520; Numerical analysis and control of bifurcation problems: (II) Bifurcation in infinite dimensions. *Int. J. Bifurcation & Chaos* **1991**, *1*, 745–772.
- [43] Doedel, E. J.; Paffenroth, R. C.; Champneys, A. R.; Fairgrieve, T. F.; Kuznetsov, Yu. A.; Sandstede, B.; Wang, X. AUTO 2000: continuation and bifurcation software for ordinary differential equations (with HomCont); Technical Report, Caltech, February 2001.
- [44] Kubíček, M.; Marek, M. *Computational Methods in Bifurcation Theory and Dissipative Structures*; New York: Springer, 1983; Marek, M.; Schreiber, I. *Chaotic Behaviour of Deterministic Dissipative Systems*; Cambridge University Press: Cambridge, 1995.
- [45] Kohout, M.; Schreiber, I.; Kubíček, M. A computational tool for nonlinear dynamical and bifurcation analysis of chemical engineering problems. *Comput. Chem. Eng.* **2002**, *26*, 517–527.
- [46] Boissonade, J.; De Kepper, P. Transitions from bistability to limit cycle oscillations— theoretical analysis and experimental evidence in an open chemical system. *J. Phys. Chem.* **1980**, *84*, 501–506.
- [47] Clarke, B. L.; Jiang, W. M. Method for deriving Hopf and saddle-node bifurcation hyper-surfaces and application to a model of the Belousov–Zhabotinskii system. *J. Chem. Phys.* **1993**, *99*, 4464–4478.
- [48] Hynne, F.; Sørensen, P. G.; Nielsen, K. Quenching of chemical oscillations: general theory. *J. Chem. Phys.* **1990**, *92*, 1747–1757; Sørensen, P. G.; Hynne, F.; Nielsen, K. Characteristic modes of oscillatory chemical reactions. *J. Chem. Phys.* **1990**, *92*, 4778–4785.
- [49] Eiswirth, M.; Bürger, J.; Strasser, P.; Ertl, G. Oscillating Langmuir–Hinshelwood mechanisms. *J. Phys. Chem.* **1996**, *100*, 19118–19123.
- [50] Eiswirth, M. Unpublished results.
- [51] Degn, H.; Olsen, L. F.; Perram, J. Bistability, oscillation, and chaos in an enzyme reaction. *Ann. N. Y. Acad. Sci.* **1979**, *316*, 623–637.
- [52] Olsen, L. F.; Hofman Frisch, L.-L.; Schaffer, W. M. In *Chaotic Hierarchy*, Baier, G. Klein, M., Eds.; World Scientific: Singapore, 1991; p 299.
- [53] Noszticzius, Z.; McCormick, W. D.; Swinney, H. L. Use of bifurcation diagrams as fingerprints of chemical mechanisms. *J. Phys. Chem.* **1989**, *93*, 2796–2800.
- [54] Olsen, R. J.; Epstein, I. R. Bifurcation analysis of chemical reaction mechanisms. 1. Steady-state bifurcation structure. *J. Chem. Phys.* **1991**, *94*, 3083–3095; Bifurcation analysis of chemical reaction mechanisms. 2. Hopf bifurcation analysis. *J. Chem. Phys.* **1993**, *98*, 2805–2822.
- [55] Ringland, J. Rapid reconnaissance of a model of a chemical oscillator by numerical continuation of a bifurcation feature of codimension-2. *J. Chem. Phys.* **1991**, *95*, 555–562.
- [56] Hynne, F.; Sørensen, P. Quenching of chemical oscillations. *J. Phys. Chem.* **1987**, *91*, 6573–6575; Sørensen, P.; Hynne, F. Amplitudes and phases of small-amplitude Belousov–Zhabotinskii oscillations derived from quenching experiments. *J. Phys. Chem.* **1989**, *93*, 5467–5474.
- [57] Hynne, F.; Sørensen, P.; Møller, T. Current and eigenvector analyses of chemical reaction networks at Hopf bifurcations. *J. Chem. Phys.* **1992**, *98*, 211–218; Complete optimization of models of the Belousov–Zhabotinsky reaction at a Hopf bifurcation. *J. Chem. Phys.* **1992**, *98*, 219–230.
- [58] Winfree, A. T. *The Geometry of Biological Time*; Springer-Verlag: New York, 1980.
- [59] Ruoff, P.; Forsterling, H.-D.; Gyorgyi, L.; Noyes, R. M. Bromous acid perturbations in the Belousov–Zhabotinsky reaction—experiments and model calculations of phase response curves. *J. Phys. Chem.* **1991**, *95*, 9314–9320.
- [60] Dolník, M.; Schreiber, I.; Marek, M. Dynamic regimes in a periodically forced reaction cell with oscillatory chemical reaction. *Physica D* **1986**, *21*, 78–92.
- [61] Davies, M. L.; Schreiber, I.; Scott, S. K. Dynamical behaviour of the Belousov–Zhabotinsky reaction in a fed-batch reactor. *Chem. Eng. Sci.* **2004**, *59*, 139–148.
- [62] Dolník, M.; Marek, M.; Epstein, I. R. Resonances in periodically forced excitable systems. *J. Phys. Chem.* **1992**, *96*, 3218–3224.
- [63] Vance, W.; Ross, J. Entrainment, phase resetting, and quenching of chemical oscillations. *J. Chem. Phys.* **1995**, *103*, 2472–2481.

- [64] Millet, I.; Vance, W.; Ross, J. Measurements of phase response in an oscillatory reaction and deduction of components of the adjoint eigenvector. *J. Phys. Chem.* **1999**, *103*, 8252–8256.
- [65] Loud, W. S. Phase shift and locking-in regions. *Q. J. Appl. Math.* **1967**, *25*, 222.
- [66] Hjelmfelt, A.; Ross, J. Experimental stabilization of unstable steady states in oscillatory and excitable systems. *J. Phys. Chem.* **1994**, *98*, 1176–1179.
- [67] Petrov, V.; Metens, S.; Borckmans, P.; Dewel, G.; Showalter, K. Tracking unstable Turing patterns through mixed-mode spatiotemporal chaos. *Phys. Rev. Lett.* **1995**, *75*, 2895–2898.
- [68] Petrov, V.; Mihaliuk, E.; Scott, S. K.; Showalter, K. Stabilizing and characterizing unstable states in high-dimensional systems from time series. *Phys. Rev. E* **1995**, *51*, 3988–3996.
- [69] Mihaliuk, E.; Skødt, H.; Hynne, F.; Sørensen, P.-G.; Showalter, K. Normal modes for chemical reactions from time series analysis. *J. Phys. Chem. A* **1999**, *103*, 8246–8251.
- [70] Nakamura, S.; Yokota, K.; Yamazaki, I. Sustained oscillations in a lactoperoxidase, NADPH and O<sub>2</sub> system. *Nature* **1969**, *222*, 794.
- [71] Degn, H. Bistability caused by substrate inhibition of peroxidase in an open reaction system. *Nature* **1968**, *217*, 1047–1050.
- [72] Olsen, L. F.; Degn, H. Chaos in an enzyme reaction. *Nature* **1977**, *267*, 177–178.
- [73] Samples, M. S.; Hung, Y.-F.; Ross, J. Further experimental studies on the horseradish peroxidase oxidase reaction. *J. Phys. Chem.* **1992**, *96*, 7338–7342.
- [74] Steinmetz, C. G.; Geest, T.; Larter, R. Universality in the peroxidase oxidase reaction—period doublings, chaos, period 3, and unstable limit cycles. *J. Phys. Chem.* **1993**, *97*, 5649–5653.
- [75] Olsen, L. F.; Kummer, U.; Kindzelskii, A. L.; Petty, H. R. A model of the oscillatory metabolism of activated neutrophils. *Biophys. J.* **2003**, *84*, 69–81.
- [76] Scheeline, A.; Olson, D. L.; Williksen, E. P.; Horras, G. A.; Klein, M. L.; Larter, R. The peroxidase–oxidase oscillator and its constituent chemistries. *Chem. Rev.* **1997**, *97*, 739–756.
- [77] Aguda, B.; Clarke, B. L. Bistability in chemical reaction networks—theory and application to the peroxidase–oxidase reaction. *J. Chem. Phys.* **1987**, *87*, 3461–3470.
- [78] Hung, Y.-F.; Schreiber, I.; Ross, J. New reaction-mechanism for the oscillatory peroxidase oxidase reaction and comparison with experiments. *J. Phys. Chem.* **1995**, *99*, 1980–1987.
- [79] Dateo, C. E.; Orban, M.; De Kepper, P.; Epstein, I. R. Systematic design of chemical oscillators. 5. Bistability and oscillations in the autocatalytic chlorite–iodide reaction in a stirred flow reactor. *J. Am. Chem. Soc.* **1982**, *104*, 504–509.
- [80] De Kepper, P.; Boissonade, J.; Epstein, I. R. Chlorite–iodide reaction—a versatile system for the study of nonlinear dynamic behavior. *J. Phys. Chem.* **1990**, *94*, 6525–6536.
- [81] Luo, Y.; Epstein, I. R. Systematic design of chemical oscillators. 38. Stirring and premixing effects in the oscillatory chlorite–iodide reaction. *J. Chem. Phys.* **1986**, *85*, 5733–5740.
- [82] Schilling, C. H.; Schuster, S.; Palsson, B. O.; Heinrich, R. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* **1999**, *15*, 296–303.

## Lifetime and Transit Time Distributions and Response Experiments in Chemical Kinetics

---

We discussed some aspects of the responses of chemical systems, linear or nonlinear, to perturbations on several earlier occasions. The first was the responses of the chemical species in a reaction mechanism (a network) in a nonequilibrium stable stationary state to a pulse in concentration of one species. We referred to this approach as the “pulse method” (see chapter 5 for theory and chapter 6 for experiments). Second, we studied the time series of the responses of concentrations to repeated random perturbations, the formulation of correlation functions from such measurements, and the construction of the correlation metric (see chapter 7 for theory and chapter 8 for experiments). Third, in the investigation of oscillatory chemical reactions we showed that the responses of a chemical system in a stable stationary state close to a Hopf bifurcation are related to the category of the oscillatory reaction and to the role of the essential species in the system (see chapter 11 for theory and experiments). In each of these cases the responses yield important information about the reaction pathway and the reaction mechanism.

In this chapter we focus on the design of simple types of response experiments that make it possible to extract mechanistic and kinetic information from complex nonlinear reaction systems. The main idea is to use “neutral” labeled compounds (tracers), which have the same kinetic and transport properties as the unlabeled compounds. In our previous work [1–12] we have shown that by using neutral tracers a class of response experiments can be described by linear response laws, even though the underlying kinetic equations are highly nonlinear. The linear response is not the result of a linearization procedure, but it is due to the use of neutral tracers. As a result the response is linear even for large perturbations, making it possible to investigate global nonlinear kinetics by making use of linear mathematical techniques. Moreover, the susceptibility functions from the response law are related to the probability densities of the lifetimes and transit

times of the various chemical species, making it easy to establish a connection between the response data and the mechanism and kinetics of the process.

The structure of this chapter is as follows. We begin by discussing the lifetime distribution of a chemical species in a complex chemical system and show that simple response experiments involving neutral tracers make it possible to evaluate the lifetime distributions from measurements. We extend the results for disordered kinetics and illustrate the approach by extracting kinetic information from a complex desorption experiment reported in the literature. We then consider more complicated response experiments involving multiple reaction pathways connecting various reaction intermediates and present methods for extracting connectivity, mechanistic, and kinetic information from experimental data. We pay special attention to the effect of different experimental as well as numerical errors on the accuracy of the results. We consider a generalization of the neutral response approach to inhomogeneous, space-dependent systems and illustrate the technique by an application in population genetics.

## 12.1 Lifetime Distributions of Chemical Species

In chemical kinetics the lifetime of an active intermediate is an important variable which characterizes the time scale of the process in which the intermediate is involved. Different techniques have been suggested for the direct experimental evaluation of this parameter; they are used extensively for the analysis of chain reactions [13] such as radical polymerization [14]. These techniques are based on the assumption that the mean lifetime of an active intermediate  $X_i$  can be evaluated as the ratio between its concentration  $x_i$  and its rate of disappearance  $W_i^-$ :

$$\langle \tau \rangle_{x_i} = x_i / W_i^- \quad (12.1)$$

In general the lifetime of an active intermediate is the time interval that elapses from the generation of the intermediate up to its disappearance. Equation (12.1) is valid only for the particular case of stationary processes for which both  $x_i$  and  $W_i^-$  are time-independent. For a time-dependent process, eq. (12.1) is no longer valid. An alternative theory has been suggested for the study of time-dependent systems, based on the use of age-dependent balance equations [1–12]. If the kinetics of a reaction system is known, this theory provides analytical expressions for the probability distributions of the lifetime of the different individuals in the system. Although the lifetime is a random variable obeying probability laws, its statistical properties can be evaluated by using deterministic kinetic equations. This type of approach is usually referred to in the literature as “semi-stochastic” [15]. In the following we give a short summary of the semi-stochastic theory of lifetime distributions.

We consider an open isothermal system made up of many active intermediates  $X_i$ ,  $i = 1, 2, \dots$ , as well as many stable species  $A_i$ ,  $i = 1, 2, \dots$ . The deterministic evolution equations for the concentrations  $x_i$ ,  $i = 1, 2, \dots$ , of the active intermediates  $X_i$ ,  $i = 1, 2, \dots$ , are given by

$$dx_i(t)/dt = W_i^+(\mathbf{a}(t), \mathbf{x}(t)) - W_i^-(\mathbf{a}(t), \mathbf{x}(t)), \quad i = 1, 2, \dots \quad (12.2)$$

where  $W_i^\pm(\mathbf{x})$ ,  $i = 1, 2, \dots$ , are the rates of generation and consumption of the different intermediates, respectively,  $x_i(t)$  and  $a_i(t)$  are the concentrations of the active

intermediates and stable species, respectively, and  $\mathbf{x}(t) = (x_i(t))$  and  $\mathbf{a}(t) = (a_i(t))$  are the corresponding concentration vectors. A similar set of kinetic equations can be written for the concentrations  $a_i(t)$  of the stable species; in some cases we can assume that the concentration of stable species is kept constant by interaction with a set of buffers.

We denote by  $\eta_i(\tau; t) d\tau$  the density of particles of type  $X_i$  at time  $t$  with a lifetime in the system between  $\tau$  and  $\tau + d\tau$ . We have

$$\int_0^{\infty} \eta_i(\tau; t) d\tau = x_i(t) \quad (12.3)$$

For evaluating these density functions we notice that the ratio

$$\lambda_i^-(t) dt = [W_i^-(\mathbf{x}(t)) / x_i(t)] dt \quad (12.4)$$

can be interpreted as the infinitesimal probability that a particle  $X_i$  disappears in a time interval  $t, t + dt$ ; the complementary probability, that a particle  $X_i$  survives in a time interval  $t, t + dt$ , is given by  $1 - \lambda_i^-(t) dt$ . By taking eqs. (12.4) into account we can derive the following evolution equations for  $\eta_i(\tau; t) d\tau$ :

$$\left( \frac{\partial}{\partial t} + \frac{\partial}{\partial \tau} \right) \eta_i(\tau; t) = - \left[ \frac{W_i^-(\mathbf{x}(t))}{x_i(t)} \right] \eta_i(\tau; t), \quad i = 1, 2, \dots \quad (12.5)$$

$$\eta_i(\tau = 0; t) = W_i^+(\mathbf{x}(t)), \quad i = 1, 2, \dots \quad (12.6)$$

Equation (12.5) is derived by writing a balance equation for the number of particles  $X_i$  and noticing that, as the time increases from  $t$  to  $t + \Delta t$ , the lifetime also increases by the same amount  $\Delta t$ , from  $\tau$  to  $\tau + \Delta t$ . For a small time interval  $\Delta t$  we have

$$\eta_i(\tau + \Delta t; t + \Delta t) = \eta_i(\tau; t) + [(\partial \eta_i(\tau; t) / \partial t) + (\partial \eta_i(\tau; t) / \partial \tau)] \Delta t + O((\Delta t)^2)$$

which leads to eqs. (12.5).

We denote by

$$\xi_i(\tau; t) d\tau \quad \text{with} \quad \int_0^{\infty} \xi_i(\tau; t) d\tau = 1 \quad (12.7)$$

the lifetime probability density of active intermediates  $X_i$  at time  $t$ . This probability density can be expressed in terms of the density function  $\eta_i(\tau; t) d\tau$  as

$$\xi_i(\tau; t) d\tau = \frac{\eta_i(\tau; t) d\tau}{\int_0^{\infty} \eta_i(\tau; t) d\tau} = \frac{\eta_i(\tau; t) d\tau}{x_i(t)} \quad (12.8)$$

We eliminate the density functions  $\eta_i(\tau; t) d\tau$  and the reaction rates  $dx_i(t)/dt$  from eqs. (12.2)–(12.8), resulting in a system of evolution equations for the lifetime probability densities  $\xi_i(\tau; t) d\tau$ :

$$\left( \frac{\partial}{\partial t} + \frac{\partial}{\partial \tau} \right) \xi_i(\tau; t) = -\lambda_i^+(t) \xi_i(\tau; t), \quad i = 1, 2, \dots \quad (12.9)$$

$$\xi_i(\tau = 0; t) = \lambda_i^+(t), \quad i = 1, 2, \dots \quad (12.10)$$

where

$$\lambda_i^+(t) dt = [W_i^+(\mathbf{x}(t))/x_i(t)] dt \quad (12.11)$$

is the probability that a new  $X_i$  intermediate is generated at a time between  $t$  and  $t + dt$ . An analytical solution of eqs. (12.9)–(12.10) can be derived by using the method of characteristics [16]. The solutions corresponding to the initial conditions,  $\xi_i(\tau; t = t_0) = \xi_i^0(\tau)$ ,  $i = 1, 2, \dots$ , are given by

$$\begin{aligned} \xi_i(\tau; t) = & h(t - t_0 - \tau) \lambda_i^+(t - \tau) \exp \left[ - \int_{t-\tau}^t \lambda_i^+(t') dt' \right] \\ & + h(\tau - t + t_0) \xi_i^0(\tau - t + t_0) \exp \left[ - \int_{t_0}^t \lambda_i^+(t') dt' \right], \quad i = 1, 2, \dots \end{aligned} \quad (12.12)$$

where  $h(x)$  is Heaviside's step function. In particular, for a stationary state we have  $\lambda_i^+, \lambda_i^-$  and  $W_i^+ = W_i^-$  independent of time, and eqs. (12.12) become

$$\xi_i(\tau) = \frac{W_i^+}{x_i} \exp \left( - \frac{x_i \tau}{W_i^+} \right) = \frac{W_i^-}{x_i} \exp \left( - \frac{x_i \tau}{W_i^-} \right) \quad (12.13)$$

and the average lifetimes are given by eq. (12.1):

$$\langle \tau \rangle_{x_i} = \int_0^\infty \tau \xi_i(\tau) d\tau = x_i / W_i^- \quad (12.14)$$

The theory can be extended to reaction systems with disordered kinetics, for which the rate coefficients are random. There are two different cases: (1) static disorder, for which the rate coefficients are random variables selected from certain probability laws; and (2) dynamic disorder, for which the rate coefficients are random functions of time. For details see [5]. Here we give the expression of the average probability density of the lifetime for stationary systems with static disorder:

$$\langle \xi_i(\tau) \rangle = \int_0^\infty \frac{W_i^-(\mathbf{k})}{x_i(\mathbf{k})} \exp \left( - \frac{x_i(\mathbf{k}) \tau}{W_i^-(\mathbf{k})} \right) p(\mathbf{k}) d\mathbf{k} \quad (12.15)$$

where  $p(\mathbf{k}) d\mathbf{k}$  is the probability density of the vector  $\mathbf{k}$  of the rate coefficients. Equation (12.15) will be used later for the theoretical analysis of the desorption response experiments of Drazer and Zanette [17].

## 12.2 Response Experiments and Lifetime Distributions

We consider that the chemical process studied takes place in an isothermal, continuous-flow stirred tank reactor (CSTR) operated with a constant total flow rate and with constant input concentrations. We assume that at least one of the chemicals entering the system is available in two different forms, unlabeled and labeled, respectively, and that the kinetic isotope effect can be neglected, that is, the kinetic parameters are the same for the labeled



and unlabeled compounds. Although the total input concentrations of chemicals are kept constant, we assume that it is possible to control the proportion of labeled compounds entering the system. Under these circumstances the overall kinetic behavior of the system, expressed in terms of total concentrations, is not affected by the introduction of tracer in the system. Only the proportions of labeled compounds in the reactor depend on the amounts of tracers introduced in the system. By measuring the concentrations of labeled chemicals in the reactor it is possible to evaluate the distributions of lifetimes of different individuals, as we now show.

For a given reaction intermediate, which may exist in an unlabeled form  $X_i$  as well as in a labeled form,  $X_i^\#$ , we assume that it is possible to measure the time dependence of the following fractions:

1. The ratio  $\alpha_i(t)$  between the rate  $W_i^{+\#}[\mathbf{x}(t)]$  of generation of the labeled compound  $X_i^\#$  and the total rate of generation  $W_i^+[\mathbf{x}(t)]$  of the two forms,  $X_i$  and  $X_i^\#$ , labeled and unlabeled:

$$\alpha_i(t) = W_i^{+\#}[\mathbf{x}(t)]/W_i^+[\mathbf{x}(t)] \quad (12.16)$$

2. The proportion  $\beta_i(t)$  of the active intermediate  $X_i^\#$ :

$$\beta_i(t) = x_i^\#(t)/x_i(t) \quad (12.17)$$

where  $x_i^\#(t)$  and  $x_i(t)$  are the concentrations of the labeled active intermediate and the total concentration of the active intermediate, labeled and unlabeled.

In order to extract the information about the lifetime distributions from these two functions,  $\alpha_i(t)$  and  $\beta_i(t)$ , we derive an evolution equation for the concentration  $x_i^\#(t)$  of labeled intermediate in the system:

$$\frac{d}{dt}x_i^\#(t) = W_i^{+\#}[\mathbf{x}(t)] - W_i^-[\mathbf{x}(t)] \frac{x_i^\#(t)}{x_i(t)} \quad (12.18)$$

Balance equations of this type were introduced in chemical kinetics by Neiman and Gál [18]. Equations (12.18) must be supplemented by the kinetic equations (12.2):

$$\frac{d}{dt}x_i(t) = W_i^+[\mathbf{x}(t)] - W_i^-[\mathbf{x}(t)] \quad (12.19)$$

For simplicity, in eqs. (12.19) the dependence of the reaction rates on the concentrations of stable species is not displayed explicitly.

By combining eqs. (12.16)–(12.19) we can derive the following relationship between  $\alpha_i(t)$  and  $\beta_i(t)$ :

$$\frac{d}{dt}\beta_i(t) + \beta_i(t) \frac{W_i^+[\mathbf{x}(t)]}{x_i(t)} = \alpha_i(t) \frac{W_i^+[\mathbf{x}(t)]}{x_i(t)} \quad (12.20)$$

These are differential equations in  $\beta_i(t)$ . By considering the initial values  $\beta_i(t_0)$ , their solutions are

$$\begin{aligned} \beta_i(t) = & \exp \left[ - \int_{t_0}^t \frac{W_i^+[\mathbf{x}(t')]}{x_i(t')} dt' \right] \beta_i(t_0) \\ & + \int_0^{t-t_0} \frac{W_i^+[\mathbf{x}(t-x)]}{x_i(t-x)} \exp \left[ - \int_{t-x}^t \frac{W_i^+[\mathbf{x}(t')]}{x_i(t')} dt' \right] \alpha_i(t-x) dx \quad (12.21) \end{aligned}$$

Equations (12.21) contain transient terms depending on the initial conditions  $\beta_i(t_0)$ ; similar transient contributions, depending on the initial conditions  $\xi_i(\tau, t_0)$ , exist in eqs. (12.12) for the lifetime distributions. In both cases, for large times the contribution of these transient terms is negligible. For simplicity we push the initial times in eqs. (12.12) and (12.21) to minus infinity,  $t_0 \rightarrow -\infty$ , and compare the resulting equations; under these circumstances the response functions  $\beta_i(t)$  can be expressed in terms of the excitation functions  $\alpha_i(t)$  and the probability densities of the lifetimes attached to the different active intermediates  $\xi_i(\tau, t)$ :

$$\beta_i(t) = \int_0^\infty \xi_i(\tau; t) \alpha_i(t - \tau) d\tau \quad (12.22)$$

By requiring that the initial conditions in eqs. (12.12) and (12.21) satisfy the law of mass conservation, it can be shown that eqs. (12.22) are valid for any initial times. Equations (12.22) are linear superposition laws, which can be used for evaluating the lifetime distributions  $\xi_i(\tau, t)$  of the active intermediates from response experiments. For nonstationary regimes, it is necessary to carry out repeated response experiments, starting at different initial times. By solving the integral equations (12.22) it is possible to compute the time and space dependence of lifetime distributions. Frequency techniques can also be used; for nonstationary regimes they also involve repeating the response experiments.

For stationary regimes the lifetime distributions  $\xi_i(\tau, t) = \xi_i(\tau)$  depend only on the lifetime and eqs. (12.22) become convolution equations:

$$\beta_i(t) = \xi_i(t) \otimes \alpha_i(t) \quad (12.23)$$

where  $\otimes$  denotes the temporal convolution product. In this case a single response experiment, either in real time or in the frequency domain, is enough for evaluating the lifetime distributions.

The response law (12.22) can be extended to disordered kinetics with static or dynamic disorder [5]. For stationary systems with static disorder the response law is

$$\langle \beta_i(t) \rangle = \int_0^\infty \langle \xi_i(\tau) \rangle \alpha_i(t - \tau) d\tau \quad (12.24)$$

where  $\langle \dots \rangle$  denotes a static average with respect to all possible values of the rate coefficients. For more general cases, see [5].

The lifetime distributions contain kinetic information. For illustrating the usefulness of our approach we consider the experiments of Drazer and Zanette on disordered

desorption kinetics [17]. These authors used nonconsolidated packings of relatively uniform, spherical, activated carbon grains obtained from apricot pits, with an average radius of  $d = (0.13 \pm 0.01)$  cm. The carbon grains were packed in a 30 cm high, 2.54 cm inner diameter cylinder. In the experiments, the porous medium was initially filled up with aqueous 0.1 M NaI solution tagged with  $I^{131}$ . The authors performed measurements of tracer adsorption and dispersion, in which a stepwise variation of the concentration of  $I^{131}$  is induced at time  $t = 0$  and kept constant thereafter. Two different types of experiments, each with a total constant flow rate, were carried out, by using different displacing fluids. In the first set of experiments the system was flushed with an untagged NaI solution having the same concentration as the initial, labeled, solution (exchange experiments), whereas in the second set of experiments the untagged NaI solution was replaced by distilled water (desorption experiments).

The results of the exchange experiments show that the replacement of the radioactive isotope  $I^{131}$  obeys a classical adsorption–desorption and dispersion mechanism commonly encountered in chemical engineering. In this case the observed concentration profiles can be reproduced theoretically by using a classical reaction–convection–diffusion equation. The time dependence of the concentration of  $I^{131}$  is described by a set of kinetic curves, which decay very quickly to zero. In the case of desorption experiments the concentration profiles  $I^{131}$  also decrease to zero for large times, but this decrease is much slower than in the case of exchange experiments. In the case of desorption for large times the concentration profiles converge toward a long time tail of the negative power law type,  $C(t) \sim t^{-\mu}$ ,  $t \gg 0$ , characterized by a fractal exponent  $\mu = 0.63$ . Such long tails cannot be explained by assuming a classical reaction–diffusion mechanism.

The experimental results of the authors suggest that the displacement of the radioactive isotope  $I^{131}$  involves a very slow, dispersive (Montroll) diffusion process [19]. The qualitative physical picture suggested by the authors is the following: the motion of an atom of the radioactive isotope along the column can be represented by a hopping mechanism involving a succession of desorption and readsorption processes, which is basically a random walk in continuous time (CTRW [19]). We developed a transport theory [7] for the Drazer–Zanette experiment which shows that the combination of hydrodynamic and rate components leads to a linear response law and that for desorption experiments the probability density of the lifetime is proportional to the concentration profile, where the time is replaced by the lifetime,  $\xi(t) \sim C(\tau) \sim \tau^{-\mu}$ , a result that is consistent with the Drazer and Zanette interpretation of their experiment. In another publication, we focused on the chemical interpretation of their result and suggested a kinetic model for their experiment [6].

Following Drazer and Zanette, we assume that, although the whole system is not at equilibrium, in different local areas of the column there is a local equilibrium, We invoke the homotatic patch approximation in heterogeneous catalysis [20], according to which the overall adsorption isotherm is made up of a contribution of local Langmuir adsorption isotherms:

$$\theta_{\text{local}} = \frac{K_{\text{local}}^{\text{eq.}} C}{1 + K_{\text{local}}^{\text{eq.}} C} \quad (12.25)$$

where  $\theta_{\text{local}}$  is the local coverage of the surface,  $C$  is the local concentration of the chemical in the solution, and

$$K_{\text{local}}^{\text{eq.}}(T) = k_{\text{local}}^+ / k_{\text{local}}^- = K_{\text{local}}^{\text{eq.}}(\infty) \exp \left[ -\frac{\Delta H}{k_B T} \right] \quad (12.26)$$

is the local equilibrium constant of the adsorption process.

$$k_{\text{local}}^{\pm}(T) = k_{\text{local}}^{\pm}(\infty) \exp \left[ -\frac{E^{\pm}}{k_B T} \right] \quad (12.27)$$

are local adsorption and desorption rates, respectively,  $E^{\pm}$  are activation energies, and

$$\Delta H = E^+ - E^- = -\varepsilon \quad (12.28)$$

is the local change of the enthalpy during the adsorption process,  $\varepsilon$  is the adsorption heat,  $k_B$  is Boltzmann's constant, and  $T$  is the temperature of the system. According to the patch approximation, different patches of the surface are characterized by different adsorption enthalpies and their statistical distribution can be described in terms of a probability density:

$$p(\varepsilon) d\varepsilon \quad \text{with} \quad \int p(\varepsilon) d\varepsilon = 1 \quad (12.29)$$

The total adsorption isotherm can be expressed as an average of the local isotherm (12.25) with respect to the adsorption heat:

$$\theta = \int \theta_{\text{local}}(\varepsilon) p(\varepsilon) d\varepsilon = \int \frac{K_{\text{local}}^{\text{eq.}}(\varepsilon, T) C}{1 + K_{\text{local}}^{\text{eq.}}(\varepsilon, T) C} p(\varepsilon) d\varepsilon \quad (12.30)$$

In addition, we assume that the activation energies  $E^{\pm}$  of the adsorption–desorption processes and the local adsorption heat,  $\varepsilon = -\Delta H$ , are related to each other by means of the linear relations:

$$E_{\pm} = E_{\pm}^0 + \beta_{\pm} \Delta H = E_{\pm}^0 - \beta_{\pm} \varepsilon \quad (12.31)$$

where  $\beta_{\pm}$  are proportionality (scaling) coefficients. The existence of the linear relationships (12.31) for adsorption kinetics has been well documented in the literature of heterogeneous catalysis, both experimentally and theoretically [21]. The relations (12.31) are referred to as Polanyi relations or “linear free energy relations,” even though they involve the adsorption enthalpy rather than the Gibbs or Helmholtz free energy. They are equivalent to linear free energy relations only if the entropy factors of the adsorption–desorption process, both thermodynamic and kinetic, are the same for all active sites on the surface (eq. (12.21)). From eqs. (12.28) and (12.31) it follows that the scaling coefficients  $\beta_{\pm}$  are related to each other by means of the relationship

$$\beta_+ = \beta_- + 1 \quad (12.32)$$

An important consequence of the linear energy relations (12.31) is that the rate coefficients  $k_{\text{local}}^{\pm}$  are deterministic functions of the adsorption heat  $\varepsilon = -\Delta H$ :

$$k_{\text{local}}^{\pm}(T) = k_{\text{local}}^{\pm}(\infty) \{\exp [\varepsilon / k_B T]\}^{\beta_{\pm}} \quad (12.33)$$

and thus, both for thermodynamic and kinetic variables, the global averages can be expressed in terms of the probability density  $p(\varepsilon)$  of the adsorption heat.

We formally represent the adsorption–desorption process as a chemical reaction:



and denote by  $[X] = C$  the concentration of chemical in the solution, by  $[Y(\varepsilon)] = y(\varepsilon) d\varepsilon$  the surface concentration of free sites with an adsorption heat between  $\varepsilon$  and  $\varepsilon + d\varepsilon$ , and by  $[Z(\varepsilon)] = z(\varepsilon) d\varepsilon$  the corresponding surface concentration of occupied sites. We also introduce a joint density function for the adsorption heat and the trapping time:

$$\eta_Z(\tau, \varepsilon) d\varepsilon d\tau \quad \text{with} \quad z(\varepsilon) d\varepsilon = d\varepsilon \int \eta_Z(\tau, \varepsilon) d\tau \quad (12.35)$$

Here  $\eta_Z(\tau, \varepsilon) d\varepsilon d\tau$  is the surface concentration of occupied sites with an adsorption heat between  $\varepsilon$  and  $\varepsilon + d\varepsilon$  and which have trapped a molecule for a time interval between  $\tau$  and  $\tau + d\tau$ . By generalizing the theory of lifetime distributions we can derive the balance equations:

$$\left( \frac{\partial}{\partial t} + \frac{\partial}{\partial \tau} \right) \eta_Z(\tau, \varepsilon) d\varepsilon d\tau = -\eta_Z(\tau, \varepsilon) k_{\text{local}}^-(\infty) \{\exp[\varepsilon/k_B T]\}^{\beta_+-1},$$

$$\eta_Z(\tau = 0, \varepsilon) = C y(\varepsilon) k_+^0 \{\exp[-\varepsilon/k_B T]\}^{\beta_+} \quad (12.36)$$

We introduce the conditional probability density

$$\varphi(\tau|\varepsilon) = \frac{\eta_Z(\tau, \varepsilon)}{\int \eta_Z(\tau, \varepsilon) d\tau} \quad \text{with} \quad \int \varphi(\tau|\varepsilon) d\tau = 1 \quad (12.37)$$

of the trapping time corresponding to a given value of the adsorption heat. Finally, the unconditional probability density of the trapping times can be evaluated by averaging over all possible values of the adsorption heat:

$$\xi(\tau) = \int \varphi(\tau|\varepsilon) p(\varepsilon) d\varepsilon \quad \text{with} \quad \int \xi(\tau) d\tau = 1 \quad (12.38)$$

Next we use the Zeldovich–Roginskii model [22], for which the adsorption heat can take any positive value between zero and infinity and the probability density  $p(\varepsilon)$  is exponential:

$$p(\varepsilon) = (k_B T^*)^{-1} \exp[-\varepsilon/(k_B T^*)] \quad \text{with} \quad \int_0^\infty p(\varepsilon) d\varepsilon = 1, T^* > T \quad (12.39)$$

The physical meaning of the Zeldovich–Roginskii model has been discussed in detail in [22]: it corresponds to a canonical distribution “frozen” at a temperature  $T^* > T$  where  $T$  is the current temperature of the system. The value of the characteristic temperature  $T^* > T$  provides information about the history of processing the surface.

At local equilibrium the adsorption rate equals the desorption rate, resulting in:

$$C y(\varepsilon) k_{\text{local}}^+(\infty) \{\exp[\varepsilon/k_B T]\}^{\beta_+} = z(\varepsilon) k_{\text{local}}^-(\infty) \{\exp[\varepsilon/k_B T]\}^{\beta_+-1} \quad (12.40)$$

In addition the total number of sites, occupied and free, is conserved, and thus we have the balance equation:

$$y(\varepsilon) + z(\varepsilon) = u(\varepsilon) \quad (12.41)$$

where  $u(\varepsilon) d\varepsilon$  is the total number of sites, free and occupied, with an adsorption heat between  $\varepsilon$  and  $\varepsilon + d\varepsilon$ . By solving eqs. (12.40)–(12.41) we arrive at

$$z(\varepsilon) = u(\varepsilon) \frac{C K_{\text{local}}^{\text{eq.}}(\infty) \exp[\varepsilon/(k_B T)]}{1 + C K_{\text{local}}^{\text{eq.}}(\infty) \exp[\varepsilon/(k_B T)]} \quad (12.42)$$

where the local equilibrium constants  $K_{\text{local}}^{\text{eq.}}(T)$ ,  $K_{\text{local}}^{\text{eq.}}(\infty)$  for temperature  $T$  are expressed as ratios of the corresponding forward and backward rate coefficients:

$$\begin{aligned} K_{\text{local}}^{\text{eq.}}(T) &= k_{\text{local}}^+(T)/k_{\text{local}}^-(T) = K_{\text{local}}^{\text{eq.}}(\infty) \exp[-\Delta H/(k_B T)] \\ K_{\text{local}}^{\text{eq.}}(\infty) &= k_{\text{local}}^+(\infty)/k_{\text{local}}^-(\infty) \end{aligned} \quad (12.43)$$

By integrating eq. (12.42) over all possible values of the adsorption heat we arrive at

$$\theta = \frac{Z(C)}{U} = \int \frac{C K_{\text{local}}^{\text{eq.}}(\infty) \exp[-\varepsilon/(k_B T)]}{1 + C K_{\text{local}}^{\text{eq.}}(\infty) \exp[-\varepsilon/(k_B T)]} p(\varepsilon) d\varepsilon \quad (12.44)$$

where

$$Z(C) = \int z(\varepsilon) d\varepsilon \quad U = \int u(\varepsilon) d\varepsilon \quad (12.45)$$

is the total surface concentration of occupied sites, and

$$U = \int u(\varepsilon) d\varepsilon \quad (12.46)$$

is the total surface concentration of active sites, free and occupied. The probability density  $p(\varepsilon) d\varepsilon$  is expressed as a ratio between the surface concentration  $u(\varepsilon) d\varepsilon$  of active sites with the adsorption heat between  $\varepsilon$  and  $\varepsilon + d\varepsilon$ , and the total surface concentration  $U$  of active sites:

$$p(\varepsilon) d\varepsilon = u(\varepsilon) d\varepsilon / U \quad (12.47)$$

We insert eq. (12.39) into eq. (12.44) and carry out the integral. After a number of transformations we obtain

$$\begin{aligned} \theta &= (C K_{\text{local}}^{\text{eq.}}(\infty))^\alpha \frac{\pi^\alpha}{\sin(\pi\alpha)} - \alpha (C K_{\text{local}}^{\text{eq.}}(\infty))^\alpha B\left(1 - \alpha, \alpha; \frac{C K_{\text{local}}^{\text{eq.}}(\infty)}{1 + C K_{\text{local}}^{\text{eq.}}(\infty)}\right) \\ &\sim (C K_{\text{local}}^{\text{eq.}}(\infty))^\alpha \frac{\pi^\alpha}{\sin(\pi\alpha)} \quad \text{as } C \rightarrow 0 \end{aligned} \quad (12.48)$$

where

$$B(p, q) = \int_0^1 x^{q-1} (1-x)^{q-1} dx, \quad q, p > 0 \quad (12.49)$$

$$B(p, q; x) = \int_0^x x^{q-1} (1-x)^{q-1} dx, \quad q, p > 0, x \geq 0 \quad (12.50)$$

are the complete and incomplete beta functions, respectively, and  $\alpha = T/T^*$ ,  $1 > \alpha > 0$ , is a scaling exponent between zero and unity. The limit  $C \rightarrow 0$  in eq. (12.48)

corresponds to the Freundlich adsorption isotherm which describes the experimental data of Drazer and Zanette.

The stationary solution of eqs. (12.36) and (12.37) is

$$\eta_Z(\tau, \varepsilon) = u(\varepsilon) \frac{C K_{\text{local}}^{\text{eq}}(\infty) \exp[\beta_+ \varepsilon / (k_B T)]}{1 + C K_{\text{local}}^{\text{eq}}(\infty) \exp[\varepsilon / (k_B T)]} \times k_{\text{local}}^-(\infty) \exp\{-\tau k_{\text{local}}^-(\infty) \exp[(\beta_+ - 1) \varepsilon / (k_B T)]\} \quad (12.51)$$

Finally we obtain the following expression for the probability density of the lifetime:

$$\xi(\tau) = \alpha \tau^{-\left(\frac{1+\alpha}{1-\beta_+}\right)} (k_{\text{local}}^-(\infty))^{-\left(\frac{\alpha+\beta_+}{1-\beta_+}\right)} \gamma\left(\frac{1+\alpha}{1-\beta_+}, \tau k_{\text{local}}^-(\infty)\right) \\ \sim \alpha (k_{\text{local}}^-(\infty))^{-\left(\frac{\alpha+\beta_+}{1-\beta_+}\right)} \Gamma\left(\frac{1+\alpha}{1-\beta_+}\right) \tau^{-\alpha_{\text{eff}}} \quad \text{as } \tau \rightarrow \infty \quad (12.52)$$

which is also compatible with the experimental results of Drazer and Zanette. Here

$$\alpha_{\text{eff.}} = (1 + \alpha) / (1 - \beta_+) \quad (12.53)$$

is an effective scaling exponent and

$$\gamma(a, x) = \int_0^x t^{a-1} \exp(-t) dt, \quad \Gamma(a) = \int_0^\infty t^{a-1} \exp(-t) dt \quad \text{with } a > 0, x \geq 0 \quad (12.54)$$

is the incomplete gamma function. In eq. (12.52), for physical consistency, we have to assume that

$$\frac{1 - \alpha}{2} > \beta_+ > -\alpha \quad (12.55)$$

These restrictions must be introduced in order to ensure the nonnegativity and the normalization to unity of the probability density of trapping times. We shall see later that restrictions (12.55) are fulfilled by the experimental data reported by Drazer and Zanette.

The desorption experiments presented in [17] show that

$$\alpha = \alpha_{\text{eff.}} = 0.63 \quad (12.56)$$

By applying eqs. (12.4) and (12.53)–(12.54) we can evaluate the numerical values of the scaling exponents  $\beta_\pm$  entering the linear energy relations (12.31):

$$\beta_+ = 0, \quad \beta_- = -1 \quad (12.57)$$

and thus

$$E_+ = E_+^0 = E_-^0 = \text{constant} \quad (12.58)$$

$$E_- = E_-^0 + \varepsilon \quad (12.59)$$

It follows that the adsorption rate is constant for all adsorption sites, irrespective of the value of the adsorption heat. The desorption rate depends on the adsorption heat in a simple way: up to a constant additive factor, the activation energy of the desorption process is equal in magnitude to the adsorption heat of the site considered. The activation energy profiles for the adsorption–desorption process have the form represented schematically

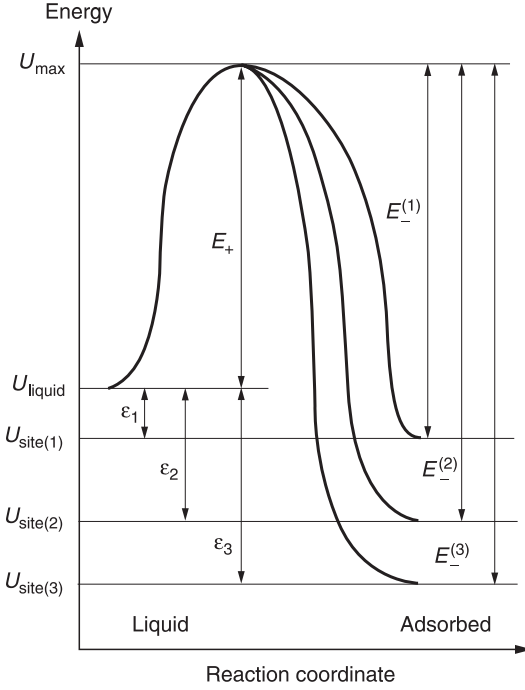


Fig. 12.1 Schematic representation of the activation energy profiles for the adsorption-desorption process studied in the experiments of Drazer and Zanette [17]. The different activation energy profiles correspond to adsorption sites characterized by different adsorption heats. All energy profiles start from the same value  $U_{\text{liquid}}$ , which corresponds to the liquid phase, increase up to the same maximum value  $U_{\text{max}}$ , and then decrease to various final values  $U_{\text{site}(1)}$ ,  $U_{\text{site}(2)}$ ,  $U_{\text{site}(3)}$ , ..., corresponding to surface sites characterized by various adsorption heats. (From [6].)

in fig. 12.1, which represents different energy profiles for the adsorption-desorption process on sites characterized by different adsorption heats. All energy profiles start from the same energy value  $U_{\text{liquid}}$  corresponding to a molecule in the liquid phase, and go up to the same maximum value  $U_{\text{max}}$ . The descending parts of the activation energy profiles are different for sites with different adsorption heats. The difference between the energy  $U_{\text{site}(j)}$ , corresponding to the surface site of type  $j$ , and the characteristic energy  $U_{\text{liquid}}$  corresponding to a molecule in the liquid phase, is equal to the variation of enthalpy during the adsorption process on the site  $j$ :

$$\Delta H_j = U_{\text{site}(j)} - U_{\text{liquid}} \leq 0 \quad (12.60)$$

The corresponding adsorption heat is equal to the variation of enthalpy with changed sign:

$$\varepsilon = -\Delta H_j = U_{\text{liquid}} - U_{\text{site}(j)} \geq 0 \quad (12.61)$$

The activation energy of the adsorption process is equal to the difference between the maximum value of the energy,  $U_{\text{max}}$ , corresponding to the top of the various energy profiles, and the characteristic energy  $U_{\text{liquid}}$ , corresponding to a molecule in the liquid:

$$E_+ = U_{\text{max}} - U_{\text{liquid}} \quad (12.62)$$

Similarly, the activation energies  $E_-^{(j)}$  of the desorption processes corresponding to the various sites of the surface are given by

$$E_-^{(j)} = U_{\text{max}} - U_{\text{site}(j)} \quad (12.63)$$



We notice that, by combining eqs. (12.61)–(12.63), we recover the linear energy relation (12.59), compatible with the experimental data of Drazer and Zanette. We have

$$E_{-}^{(j)} = U_{\max} - U_{\text{site}(j)} = (U_{\max} - U_{\text{liquid}}) + (U_{\text{liquid}} - U_{\text{site}(j)}) = E_{-}^0 + \varepsilon \quad (12.64)$$

where

$$E_{-}^0 = U_{\max} - U_{\text{liquid}} = E_{+} = E_{+}^0 \quad (12.65)$$

In conclusion, the lifetime analysis of the experimental data of Drazer and Zanette provided important clues regarding the kinetics and mechanism of the desorption process. In particular, the knowledge of the fractal exponents of the adsorption isotherm and of the tail of the lifetime distribution was enough to elucidate the shape and structure of the activation energy barrier. We have also shown that the experimental data indicate that the adsorption rate and the adsorption activation energies are constant and only the desorption rate and desorption activation energy are random. The application of the method may involve detailed theoretical developments for different systems; nevertheless we expect that the results are worth the effort.

## 12.3 Transit Time Distributions in Complex Chemical Systems

Response experiments that can be interpreted in terms of lifetime distributions are limited to the study of the response of an excitation of the same species. Because of this constraint the information acquired is local and refers to a single species, which is usually part of a large reaction network. In order to obtain global information about reaction mechanisms and kinetics, we need to involve responses to excitations involving at least two species, preferably more.

We consider a complex chemical system and focus on a set of  $S$  species  $M_u$ ,  $u = 1, \dots, S$ , which can carry one or more identical molecular fragments that are unchanged during the process; in the following we refer to these species as “carriers.” For simplicity, in this section we limit ourselves to the case of isothermal, well-stirred, homogeneous systems, for which the concentrations  $c_u = c_u(t)$ ,  $u = 1, \dots, S$ , of the chemicals  $M_u$ ,  $u = 1, \dots, S$ , are space independent and depend only on time. Later on we consider the more complicated case of reaction–diffusion systems. The deterministic kinetic equations of the process can be expressed in the following form:

$$dc_u(t)/dt = \mathcal{J}_u^{+}(t) - \mathcal{J}_u^{-}(t) + \rho_u^{+}(\mathbf{c}; t) - \rho_u^{-}(\mathbf{c}; t), \quad u = 1, \dots, S \quad (12.66)$$

where  $\rho_u^{\pm}(\mathbf{c}; t)$  are the rates of formation and consumption of the species  $M_u$ , respectively,  $\mathcal{J}_u^{\pm}(t)$  are the input and output fluxes of  $M_u$ , respectively, and

$$\mathbf{c}(t) = [c_u(t)]_{u=1, \dots, S} \quad (12.67)$$

is the composition vector of the system. In general the functions  $\rho_u^{\pm}(\mathbf{c}; t)$  can be found with the mass action law or other kinetic laws. Together with the initial condition

$$\mathbf{c}(t = t_0) = \mathbf{c}_0 \quad (12.68)$$

the kinetic equations (12.67) determine the time evolution of the concentration vector.

We denote by  $z_u$  the number of fragments in the carrier species  $u$ . We use the notation

$$F_u = F(M_u), \quad u = 1, \dots, S \quad (12.69)$$

for a molecular fragment in the carrier  $M_u$ . All fragments  $F_u$  in different carriers  $M_u$  have the same structure; the label  $u$  means that a fragment belongs to a given carrier. The concentrations  $f_u(t)$ ,  $u = 1, \dots, S$ , of the fragments  $F_u$ ,  $u = 1, \dots, S$ , which belong to different carriers are given by

$$f_u(t) = z_u c_u(t), \quad u = 1, \dots, S \quad (12.70)$$

We express the kinetic equations (12.66) in terms of the concentrations  $f_u(t)$ ,  $u = 1, \dots, S$ , of the fragment  $F_u$ ,  $u = 1, \dots, S$ , in the carrier  $u$ , resulting in

$$df_u(t)/dt = J_u^+(t) - J_u^-(t) + R_u^+(\mathbf{c}; t) - R_u^-(\mathbf{c}; t), \quad u = 1, \dots, S \quad (12.71)$$

where

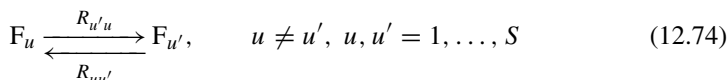
$$R_u^\pm(\mathbf{c}; t) = z_u \rho_u^\pm(\mathbf{c}; t), \quad u = 1, \dots, S \quad (12.72)$$

are the rates of formation and consumption of the fragment  $F_u$ , in the carrier  $u = 1, \dots, S$ , and

$$J_u^\pm(t) = z_u \mathcal{J}_u^\pm(t) \quad (12.73)$$

are input and output fluxes of fragment  $F_u$ , in the carrier  $u = 1, \dots, S$ , respectively.

In the system a fragment is transferred from one carrier to another. These transfer processes involving fragments  $F_u$ ,  $u = 1, \dots, S$ , among different carriers can be formally represented as:



where  $R_{u'u} = R_{u'u}(\mathbf{c}; t)$  is the rate of transport of the fragment  $F_u$  from a carrier  $M_u$  into a carrier  $M_{u'}$ . The rates  $R_{u'u}$  are related to the formation and consumption rates  $R_u^\pm(\mathbf{c}; t)$  of the fragment  $F_u$  in the carrier  $M_u$  by means of the balance equations

$$R_u^+(\mathbf{c}; t) = \sum_{u' \neq u}^S R_{uu'}(\mathbf{c}; t), R_u^-(\mathbf{c}; t) = \sum_{u' \neq u}^S R_{u'u}(\mathbf{c}; t), \quad u = 1, \dots, S \quad (12.75)$$

If the mechanism and kinetics of the process are known, then the transformation rates  $R_{u'u} = R_{u'u}(\mathbf{c}; t)$  can be evaluated by using the mass action law or other kinetic laws.

Now we consider a kinetic tracer experiment by assuming that a fraction  $\alpha_u$ ,  $u = 1, \dots, S$ , of the "in" flux  $J_u^+(t)$  of the fragment  $F_u$  is replaced by a labeled fragment  $F_u^*$ , and assume that there is no kinetic isotope effect, that is, the rates of the processes involving labeled compounds are the same as the rates of the processes involving unlabeled chemicals. We assume that the fractions  $\beta_u$ ,  $u = 1, \dots, S$ , of the labeled fragments in the output fluxes  $J_u^-(t)$ ,  $u = 1, \dots, S$ , can be measured experimentally. We are interested in answering the following questions:

1. If the mechanism and kinetics of the process are known, and if the kinetic isotope effect can be neglected, what is the general relationship between the input fractions  $\alpha_u$ ,  $u = 1, \dots, S$ , of labeled fragments and the output fractions  $\beta_u$ ,  $u = 1, \dots, S$ , of labeled fragments?

2. If the kinetics of the process is known, what are the lifetime probability densities of the fragments from different carriers present in the system and the probability densities of the time necessary for a fragment to cross the system?
3. If the kinetics and the mechanism of the process are unknown, what useful mechanistic and kinetic information can be extracted from the response experiments?

Although basically similar to the single species case presented before, the derivation of a response law for many species is more complicated. In the following, we give only the main balance equations resulting from the application of the assumption that the kinetic properties of the labeled and unlabeled species are the same (neutrality condition), and the resulting response law.

We introduce the specific rates of transport of a fragment from one carrier to another and in and out of the system:

$$\omega_{uu'}(t) = R_{uu'}(t)/f_{u'} \quad (12.76)$$

$$\Omega_u^\pm = J_u^\pm(t)/f_u \quad (12.77)$$

Here  $\omega_{uu'}(t) dt = dt R_{uu'}(t)/f_{u'}$  can be interpreted as the infinitesimal probability of transport of a fragment from the carrier  $M_{u'}$  to the carrier  $M_u$  at a time between  $t$  and  $t + dt$ ; similarly,  $\Omega_u^\pm dt = dt J_u^\pm(t)/f_u$  can be interpreted as the infinitesimal probability that a fragment in the carrier  $M_u$  enters or leaves the system at a time between  $t$  and  $t + dt$ , respectively. If the kinetic isotope effect is missing, the rate of exchange of the labeled fragments in the system can be completely expressed in terms of these infinitesimal probabilities. We assume that the time dependences of the total rates  $R_{u'u} = R_{u'u}(\mathbf{c}; t)$  and  $J_u^\pm(t) = z_u \mathcal{F}_u^\pm(t)$  attached to the total amounts of fragments from different carriers, labeled and unlabeled, are not changed during the process. We use the notations  $f_u^*(t)$ ,  $u = 1, \dots, S$ , for the concentrations of labeled fragments and  $J_u^{\pm*}(t)$ ,  $u = 1, \dots, S$ , for the input and output fluxes of labeled fragments, respectively. We use again the kinetic isotope approach in the form suggested by Neiman and Gál [18], from which we can derive the balance equations:

$$\begin{aligned} df_u^*(t)/dt &= J_u^{+*}(t) - J_u^{-*}(t) + \sum_{u' \neq u}^S [R_{uu'}^{+*}(\mathbf{c}; t) - R_{u'u}^{-*}(\mathbf{c}; t)] \\ &= J_u^{+*}(t) - \Omega_u^-(t) f_u^*(t) + \sum_{u' \neq u}^S [\omega_{uu'}(t) f_{u'}^*(t) - \omega_{u'u}(t) f_u^*(t)] \end{aligned} \quad (12.78)$$

and

$$J_u^{-*}(t) = \Omega_u^-(t) f_u^*(t) \quad (12.79)$$

Our purpose is to derive a relationship between the fractions of labeled fragments in the input fluxes:

$$\alpha_u(t) = J_u^{+*}(t)/J_u^+(t) \quad (12.80)$$

and the fractions of labeled fragments in the output fluxes:

$$\beta_u(t) = J_u^{-*}(t)/J_u^-(t) \quad (12.81)$$

From eqs. (12.78)–(12.79) we can derive the following response laws for multiple species:

$$\beta_u(t) = \sum_{u'=1}^S \int_{-\infty}^t \chi_{uu'}(t; t') \alpha_{u'}(t') dt' \quad (12.82)$$

where  $\chi_{uu'}(t; t')$  are nonnegative susceptibility functions that fulfill the normalization condition:

$$\sum_{u'=1}^S \int_{-\infty}^t \chi_{uu'}(t; t') dt' = 1 \quad (12.83)$$

Two types of expressions have been derived for the susceptibility functions  $\chi_{uu'}(t; t')$ . A first type depends on a set of direct Green functions  $G_{uu'}(t, t') = [\mathbf{G}(t, t')]_{uu'}$ , which are the solutions of the matrix differential equation

$$\frac{d}{dt} \mathbf{G}(t, t') = \mathbf{K}(t) \mathbf{G}(t, t') \quad \text{with } \mathbf{G}(t = t', t') = \mathbf{I} \quad (12.84)$$

with

$$\mathbf{K}(t) = \left[ (1 - \delta_{uu'}) \omega_{uu'}(t) - \delta_{uu'} \left[ \Omega_u^-(t) + \sum_{u'' \neq u} \omega_{u''u}(t) \right] \right]_{u, u'=1, \dots, S} \quad (12.85)$$

The expressions for the susceptibility functions are:

$$\chi_{uu'}(t; t') = \frac{J_{u'}^+(t') G_{uu'}(t; t')}{f_u(t)} \quad (12.86)$$

A second type of expressions depends on a set of adjoint Green functions  $\hat{G}_{uu'}(t, t') = [\hat{\mathbf{G}}(t, t')]_{uu'}$ , which are the solutions of the matrix differential equation

$$\frac{d}{dt} \hat{\mathbf{G}}(t, t') = \hat{\mathbf{K}}(t) \hat{\mathbf{G}}(t, t') \quad \text{with } \hat{\mathbf{G}}(t = t', t') = \mathbf{I} \quad (12.87)$$

with

$$\hat{\mathbf{K}}(t) = \left[ (1 - \delta_{uu'}) \hat{\omega}_{uu'}(t) - \delta_{uu'} \left[ \Omega_u^+(t) + \sum_{u'' \neq u} \hat{\omega}_{u''u}(t) \right] \right]_{u, u'=1, \dots, S} \quad (12.88)$$

The expressions for the susceptibility functions are

$$\chi_{uu'}(t, t') = \hat{G}_{uu'}(t; t') \Omega_{u'}^+(t') \quad (12.89)$$

The physical interpretation of the response equations (12.82) is related to a generalization of the lifetime probability densities, the transit time probability densities. We define these probability densities in terms of the following density functions:

$$\zeta_{uj}(\theta; t) d\theta \quad \text{with } \sum_j \int_0^\infty \zeta_{uj}(\theta; t) d\theta = f_u(t) \quad (12.90)$$

$\zeta_{uj}(\theta; t) d\theta$  is the concentration of fragments in the carrier  $M_u$  present in the system at time  $t$  which have entered the system in a carrier molecule of type  $j$  and have a

transit (residence) time in the system between  $\theta$  and  $\theta + d\theta$ . In terms of these density functions, we can introduce various probability densities of the transit time (see [9]). In the following we focus on only one probability of the transit time, which has a direct connection with the response law (12.82):

$$\varphi_{uu'}(\theta; t)d\theta = \frac{\zeta_{uu'}(\theta; t)d\theta}{\sum_{u'} \int_0^\infty \zeta_{uu'}(\theta; t)d\theta} = \frac{\zeta_{uu'}(\theta; t)d\theta}{f_u(t)} \quad (12.91)$$

with the normalization condition

$$\sum_{u'} \int_0^\infty \varphi_{uu'}(\theta; t)d\theta = 1 \quad (12.92)$$

$\varphi_{uu'}(\theta; t)d\theta$  is the probability that a fragment that is present in the system at time  $t$ , in a known carrier of type  $u$ , has a random transit (residence) time between  $\theta$  and  $\theta + d\theta$  and entered the system in a carrier of random type  $u'$ .

The density functions  $\zeta_{uj}(\theta; t)$  obey the following balance equations:

$$\left( \frac{\partial}{\partial t} + \frac{\partial}{\partial \theta} \right) \zeta_{uj}(\theta; t) = \sum_{u' \neq u} \omega_{uu'}(t) \zeta_{u'j}(\theta; t) - \left[ \Omega_u^-(t) + \sum_{u' \neq u} \omega_{u'u}(t) \right] \zeta_{uj}(\theta; t) \quad (12.93)$$

with the boundary conditions

$$\zeta_{uj}(\theta = 0; t) = \delta_{uj} J_u^+(t) \quad (12.94)$$

Equations (12.93)–(12.94) are similar to eqs. (12.9)–(12.10). The computation of the probability densities of the transit time is similar to the computation of lifetime probability densities. We integrate eqs. (12.93)–(12.94) along the characteristics with the initial conditions

$$\zeta_{uj}(\theta; t = t_0) = \zeta_{uj}^{(0)}(\theta) \quad (12.95)$$

and use eqs. (12.91)–(12.92). After lengthy manipulations we arrive at

$$\varphi_{uu'}(\theta; t) = \frac{G_{uu'}(t, t - \theta) J_{u'}^+(t - \theta)}{f_u(t)} = \chi_{uu'}(t; t' = t - \theta) \quad (12.96)$$

It follows that the probability density  $\varphi_{uu'}(\theta; t)$  of the transit time is equal to the susceptibility function  $\chi_{uu'}(t; t')$  evaluated for an initial time  $t' = t - \theta$ . Conversely, the susceptibility function  $\chi_{uu'}(t; t')$  is equal to the probability density  $\varphi_{uu'}(\theta; t)$  evaluated for a transit (residence) time equal to  $\theta = t - t'$ :

$$\chi_{uu'}(t; t') = \varphi_{uu'}(\theta = t - t'; t) \quad (12.97)$$

Equation (12.97) establishes a relationship between the susceptibility function  $\chi_{uu'}(t; t')$  and the probability densities  $\varphi_{uu'}(\theta; t)$  of the transit time of a fragment crossing the system.

In order to clarify the physical meaning of the transit time probability densities  $\varphi_{uu'}(\theta; t)$ , we introduce an additional random variable, the number of transport events

$q$  from one carrier to another of a given fragment present in the system. We consider a succession of  $q$  transport events of a fragment:

$$u'_0 \rightarrow u'_1 \rightarrow \dots \rightarrow u'_q \rightarrow u \quad (12.98)$$

To each transport event we attach a lifetime, for the transition  $u'_0 \rightarrow u'_1$ , the lifetime  $\tau_1$  of a fragment in the carrier  $u'_0$ , for the transition  $u'_1 \rightarrow u'_2$ , the lifetime  $\tau_2$  of a fragment in the carrier  $u'_1$ , ..., and finally the lifetime  $\tau_{q+1}$  of a fragment in the carrier  $u'_q$ . The total transit time  $\theta$  attached to the process (12.98) is simply the sum of all lifetimes attached to the different processes (12.98):

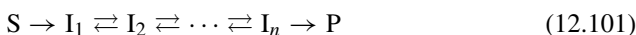
$$\theta = \tau_1 + \dots + \tau_{q+1} \quad (12.99)$$

It follows that the probability density  $\varphi_{uu'}(\theta; t)$  of the transit time  $\theta$  at time  $t$  can be expressed as a multiple average of a delta function  $\delta\left[\theta - \sum_{q'=1}^{q+1} \tau_{q'}\right]$  taken over all possible transitions  $u'_0 \rightarrow u'_1 \rightarrow \dots \rightarrow u'_q \rightarrow u$ , all possible intermediate states  $u'_1, \dots, u'_q$ , all possible lifetimes  $\tau_1, \dots, \tau_{q+1}$ , and all possible numbers  $q$  of transport events:

$$\varphi_{uu'}^{\text{tr.}}(q, \theta; t) = \left\langle \delta \left[ \theta - \sum_{q'=1}^{q+1} \tau_{q'} \right] \right\rangle \quad (12.100)$$

Thus, the probability density of transit times can be interpreted as a path sum, which is a discrete analog of a path integral, and which expresses the individual contributions of various successions of transformations of different lengths of the type  $u'_0 \rightarrow u'_1 \rightarrow \dots \rightarrow u'_q \rightarrow u$ , to the transport of a fragment within the system. Similarly, a random realization of the transit time  $\theta$  is the sum of the various lifetimes  $\tau_1, \dots, \tau_{q+1}$  of the states  $u'_1, \dots, u'_q, u$ .

Equation (12.99) looks similar to a relationship derived by Easterby [23] for a linear reaction sequence of  $n$  intermediate species (metabolites),  $I_1, \dots, I_n$ , operated in stationary conditions:



for which the average transit time  $\langle \theta \rangle$  can be expressed as a sum of average lifetimes  $\langle \tau_1 \rangle, \dots, \langle \tau_n \rangle$  corresponding to the intermediates  $I_1, \dots, I_n$ :

$$\langle \theta \rangle = \sum_{j=1}^n \langle \tau_j \rangle = \sum_{j=1}^n \frac{[I_j^{\text{st}}]}{J^{\text{st}}} \quad (12.102)$$

where  $[I_1^{\text{st}}], \dots, [I_n^{\text{st}}]$  are the stationary concentrations of the species  $I_1, \dots, I_n$ ,  $J^{\text{st}}$  is the stationary value of the reaction flux along the linear reaction sequence (12.101), and  $\langle \tau_1 \rangle, \dots, \langle \tau_n \rangle$  are the average lifetimes of the species  $I_1, \dots, I_n$ :

$$\langle \tau_j \rangle = [I_j^{\text{st}}] / J^{\text{st}}, \quad j = 1, \dots, n \quad (12.103)$$

Despite the apparently similar structure between eqs. (12.99) and (12.102), there is a fundamental difference: in eq. (12.99) all quantities are fluctuating; they are random realizations of various different random variables, which are generally correlated, whereas Easterby's equation (12.102) is a deterministic relation connecting the average values

of these random variable in the particular case of a linear reaction sequence operated in stationary conditions. Within the framework of our approach, the stochastic dependence among the various lifetimes is generally non-Markovian, and a simple relationship between the average transit time  $\langle \theta \rangle$  and the average lifetimes  $\langle \tau_1 \rangle, \dots, \langle \tau_{q+1} \rangle$  similar to Easterby's equation (12.102) does not exist.

In the particular case of a stationary regime, the stochastic dependence among the random variables attached to the different carriers becomes Markovian and our formalism becomes much simpler, and we can express the transit time corresponding to a state  $u$ ,  $\langle \theta \rangle_u$ , in terms of the lifetimes of the intermediates corresponding to the different pathways:

$$\langle \theta \rangle_u = \sum_{u'} \langle \tau \rangle_{uu'} + \sum_{q=1}^{\infty} \sum_{u'_q} \cdots \sum_{u'_0} \left( \langle \tau \rangle_{uu'_q} + \cdots + \langle \tau \rangle_{uu'_0} \right) \quad (12.104)$$

Equation (12.104) is a generalization of Easterby's equation (12.102). In the particular case of a linear reaction sequence of the type (12.101), eq. (12.104) reduces to the original Easterby relation (12.102).

In conclusion, we have introduced a "neutral" type of linear response experiment for nonlinear kinetics involving multiple reaction intermediates. We have shown that the susceptibility functions from the response equations are given by the probability densities of the transit time in the system. We have shown that a transit time is a sum of different lifetimes corresponding to different reaction pathways, and that in the particular case of a time-invariant system our definition of the transit time is consistent with Easterby's definition [23].

## 12.4 Transit Time Distributions, Linear Response, and Extracting Kinetic Information from Experimental Data

In order to extract kinetic information from tracer experiments, we express the response theorem (12.82) in terms of the probability density  $\varphi_{uu'}(\theta; t)$  (eq. (12.91)). We have

$$\beta_u(t) = \sum_{u'=1}^S \int_0^{\infty} \varphi_{uu'}(\theta; t) \alpha_{u'}(t - \theta) d\theta \quad (12.105)$$

We consider two different types of response experiments: (1) transient experiments, for which the excitation functions  $\alpha_{u'}(t)$  have the form of a unitary impulse or of a Heaviside step function; (2) frequency response experiments, for which the excitation functions  $\alpha_{u'}(t)$  are periodic.

An example of the first type of transient response experiment consists in the generation of an impulse in the input channel  $u_0$ , at different initial times  $t_0 = t_0^{(1)}, t_0^{(2)}, \dots$ :

$$\alpha_u(t) = A_{u_0} \delta_{uu_0} \delta(t - t_0) \quad (12.106)$$

where  $A_{u_0}$  is an amplitude factor. According to eq. (12.105) the response to the excitation (12.106) is

$$\beta_u(t)|_{t_0} = A_{u_0} \varphi_{uu_0}(t - t_0; t) \quad (12.107)$$

from which we obtain the following expression for the probability density  $\varphi_{uu_0}(\theta; t)$ :

$$\varphi_{uu_0}(\theta; t) = \frac{\beta_u(t)|_{t_0=t-\theta}}{A_{u_0}} \quad (12.108)$$

We notice that, in order to evaluate the probability density  $\varphi_{uu_0}(\theta; t)$ , it is necessary to carry out many response experiments, for which the excitation of the system takes place at different initial times  $t_0 = t_0^{(1)}, t_0^{(2)}, \dots$ . The method simplifies considerably if the system is operated in a stationary regime. In this case, owing to the property of temporal invariance, we have

$$\beta_u(t)|_{t_0} = \beta_u(t - t_0) \quad (12.109)$$

and (12.108) reduces to

$$\varphi_{uu_0}(\theta) = \frac{\beta_u(\theta)}{A_{u_0}}, \quad \text{independent of } t_0 \quad (12.110)$$

Here a single set of response experiments is enough for the evaluation of the probability density  $\varphi_{uu_0}(\theta; t)$  of the transit time.

The generation of an impulse with the shape of a Dirac delta function is a difficult experimental problem. It is usually easier to generate a response having the shape of a step function of the type

$$\alpha_u(t) = B_{u_0} \delta_{uu_0} h(t - t_0) \quad (12.111)$$

where  $B_{u_0}$  is another amplitude factor and  $h(t - t_0)$  is the unitary Heaviside step function. The response to the excitation (12.111) is

$$\beta_u(t)|_{t_0} = B_{u_0} \int_0^{t-t_0} \varphi_{uu_0}(x; t) dx \quad (12.112)$$

from which we obtain the following expressions for the probability density  $\varphi_{uu_0}(\theta; t)$ :

$$\varphi_{uu_0}(\theta; t) = -\frac{1}{B_{u_0}} \frac{\partial}{\partial t_0} \left[ \beta_u(t)|_{t_0} \right] \Big|_{t_0=t-\theta} \quad (12.113)$$

In particular, if the system is operated in a stationary regime, the equations become simpler:

$$\varphi_{uu_0}(\theta) = \frac{1}{B_{u_0}} \frac{\partial}{\partial \theta} \beta_u(\theta) \quad (12.114)$$

For a typical frequency response experiment the excitation function  $\alpha_{u'}(t)$  is a periodic function characterized by a single harmonic, corresponding to the carrier  $u_0$ :

$$\alpha_{u'}(t) = \left[ \widetilde{\alpha_{u_0}} + \Delta \alpha_{u_0} \exp(i\varpi t) \right] \delta_{u'u_0} \quad (12.115)$$

where  $\varpi$  is a characteristic frequency, and  $\widetilde{\alpha_{u_0}}$  and  $\Delta \alpha_{u'}$  are the temporal average and the amplitude of the excitation function, respectively. The response is given by

$$\beta_u(t) = \widetilde{\alpha_{u_0}} \Xi_{uu_0}(\varpi = 0; t) + \Delta \alpha_{u_0} \Xi_{uu_0}(\varpi; t) \exp(i\varpi t) \quad (12.116)$$



where

$$\Xi_{uu'}(\varpi; t) = \int_0^\infty \varphi_{uu'}(\theta; t) \exp(-i\varpi\theta) d\theta \quad (12.117)$$

is a frequency-dependent susceptibility function given by the one-sided Fourier transform of the transit time distribution  $\varphi_{uu'}(\theta; t)$ . The frequency response of the system, expressed by eqs. (12.116)–(12.117), can be rather complicated: for example, if the underlying kinetics of the process generates a stable limit cycle, it is made up of the superposition of two different oscillatory processes, an external oscillation given by the excitation and an internal oscillation, represented by the complex susceptibility  $\Xi_{uu'}(\varpi; t)$ , which, like the transit time probability density,  $\varphi_{uu'}(\theta; t)$ , is a periodic function of time. The interaction between these two oscillations may lead to interesting effects [2].

If the experimental data are sufficiently accurate, then from the susceptibility function  $\Xi_{uu'}(\varpi; t)$  it is possible to evaluate the transit time probability density,  $\varphi_{uu'}(\theta; t)$ , by means of numerical inverse Fourier transformation. If the experimental data are not very accurate, then it is still possible to evaluate the first two or three moments and cumulants of the transit time. Since the mean transit time is nonnegative, it follows that  $\varphi_{uu'}(\theta < 0; t) = 0$ , and thus the characteristic function of the probability density  $\varphi_{uu'}(\theta; t)$  is identical with the susceptibility function in the frequency domain,  $\Xi_{uu'}(\varpi; t)$ , experimentally accessible from eqs. (12.116)–(12.117). It follows that the moments  $\langle \theta^m(t) \rangle_u$ ,  $m = 1, 2, \dots$ , and the cumulants  $\ll \theta^m(t) \gg_u$ ,  $m = 1, 2, \dots$ , of the transit time can be evaluated by computing the derivatives of  $\Xi_{uu'}(\varpi; t)$  and  $\ln \Xi_{uu'}(\varpi; t)$  for zero frequency, respectively:

$$\langle \theta^m(t) \rangle_u = i^m \sum_{u'=1}^S \frac{d^m}{d\varpi^m} \Xi_{uu'}(\varpi = 0; t), \quad m = 1, 2, \dots \quad (12.118)$$

$$\ll \theta^m(t) \gg_u = i^m \sum_{u'=1}^S \frac{d^m}{d\varpi^m} \ln \Xi_{uu'}(\varpi = 0; t), \quad m = 1, 2, \dots \quad (12.119)$$

The qualitative and quantitative analysis of the probability densities of the transit time may lead to interesting hints concerning the reaction mechanism and the kinetics of the process. This is an important problem that needs further development. Here we present only a simple illustration based on the study of the response of a time-invariant (stationary) system.

In general, in the response law (12.105) the nonlinearity of the underlying kinetics leads to a nonstationary, time-dependent regime, for which the delay functions  $\chi_{uu'}(t, t')$  depend on two different times. In this case, the investigation of the reaction mechanisms and the kinetics is rather difficult. However, if the chemical process can be operated in a stationary regime, then the response experiment can be described by stationary time series and the delay function  $\chi_{uu'}(t, t')$  depends only on the difference,  $t - t'$ , of the entrance and exit times:

$$\chi_{uu'}(t, t') = \chi_{uu'}(t - t') = \varphi_{uu'}(\theta) \quad (12.120)$$

where  $\varphi_{uu'}(\theta)$  is the probability density of the transit time  $\theta = t - t'$ . For a stationary process the probability density of the transit time  $\varphi_{uu'}(\theta)$  can be expressed in terms

of a stationary transport matrix  $\mathbf{K} = |K_{uu'}|$ , where  $K_{uu'}$  is the rate of transport of a molecular fragment from a carrier  $u$  to a carrier  $u'$ . In this case the investigation of reaction mechanisms by means of tracer experiments is much simpler.

The qualitative examination of the shape of the response curve may lead to interesting conclusions concerning the reaction mechanism. A simple procedure consists in the examination of the shape of a log-log plot of the effective decay rate  $k_{\text{eff}}(\theta) = -\partial \ln \varphi_{uu'}(\theta) / \partial \theta$  versus the transit time,  $\log k_{\text{eff}}(\theta)$  versus  $\log \theta$ . The number of horizontal regions displayed by such a log-log plot gives a limit for the minimum number of reaction steps involved in the process. Such a method works only if the time scales corresponding to the different reaction steps have different orders of magnitude. An alternative approach, which works if the time scales of the various chemical steps are close to each other, is based on the fitting of the delay function  $\chi_{uu'}(t - t') = \varphi_{uu'}(\theta)$  by a superposition of exponential factors:

$$\varphi_{uu'}(\theta) = \sum_m A_m \exp(-\lambda_m \theta) \quad (12.121)$$

The minimum number of exponential factors in eq. (12.121) necessary for fitting the experimental data provides a lower limit for the number of chemical steps involved in the process.

Our method makes it possible to use simple linear rules for exploring complicated nonlinear systems. A simple application is the study of connectivity among various chemical species in complicated reaction networks. In the simple case of homogeneous systems with time-invariant structure, the susceptibility matrix  $\chi = [\chi_{uu'}] = \chi(\tau)$  depends only on the transit time and not on time itself. The matrix elements  $\chi_{uu'}(\tau)$  are proportional to the elements  $G_{uu'}(\tau)$  of a Green function matrix  $\mathbf{G}(\tau) = [G_{uu'}(\tau)]$ , which is the exponential of a connectivity matrix  $\mathbf{K}$ , that is,  $\mathbf{G}(\tau) = \exp[\tau \mathbf{K}]$ . It follows that from a response experiment involving a system with time-invariant structure, it is possible to evaluate the connectivity matrix,  $\mathbf{K}$ , which contains information about the relations among the different chemical species involved in the reaction mechanism. The nondiagonal elements of the matrix  $\mathbf{K} = |K_{uu'}|$  show whether in the reaction mechanism there is a direct connection between two species; in particular, if  $K_{uu'} \neq 0$ , there is a connection from the species  $u$  to the species  $u'$ ; the reverse connection, from  $u'$  to  $u$ , exists if  $K_{u'u} \neq 0$ .

Important information about the reaction kinetics and mechanisms can be obtained by investigating the concentration dependence of the fitted eigenvalues  $\lambda_m$  from (12.121). This dependence can be extremely complicated. In order to circumvent this complication, we suggest to analyze the concentration dependence of a set of the tensor invariants [24] constructed from the eigenvalues  $\lambda_m$ :

$$\mathcal{J}_w = \sum_{v_1} \cdots \sum_{v_w} \lambda_{v_1} \cdots \lambda_{v_w} \quad (12.122)$$

The reaction invariants depend on the concentrations of the different species present in the system. In most cases the dependence  $\mathcal{J}_w = \mathcal{J}_w(\text{concentrations})$  can be expressed by a polynomial, where the coefficients of the concentrations depend on the rate coefficients of the process. Starting from different assumed reaction mechanisms, we can

derive theoretical expressions for the dependence of the invariants on the concentrations. By checking the validity of these dependences for the experimental data we can test the validity of the assumed reaction mechanisms.

In conclusion, we have suggested that the linear response law and the response experiments can be applied to the study of dynamic behavior of complex chemical systems. We have shown that the response experiments make it possible to evaluate the susceptibility functions from transient as well as frequency response experiments. We have shown that the susceptibility functions bear important information about the mechanism and kinetics of complex chemical processes. We have suggested a method, based on the use of tensor invariants, which may be used for extracting information about reaction mechanism and kinetics from susceptibility measurements in time-invariant systems.

## 12.5 Errors in Response Experiments

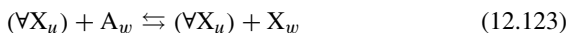
The experimental study of large chemical and biochemical systems involves simultaneous measurements of many variables, typically concentrations. Extracting kinetic information from such large amounts of data is not easy. One way of overcoming the complications due to nonlinear kinetics is to use small perturbations and use linearized kinetics for the analysis of experimental data. This is a popular approach in biochemistry and molecular biology (see for example [25]). Unfortunately, for small perturbations experimental errors are high, in the range of 20–30% or even higher, and thus the accuracy of the results obtained based on small perturbations is questionable. Compared to these methods, our approach has a big advantage; it is based on a linear response law which is not the result of a linearization procedure and is valid even for large perturbations and nonlinear kinetics.

In the following we analyze the influence of errors on our approach and its accuracy and compare the results with those obtained by using linearized kinetics. We consider a nonlinear kinetic example for which a detailed analytical study is possible. We compare that exact solution with the first-order response theory based on appropriate tracer measurements, and also compare it with the response of the linearized kinetic example. An important interest here is in the effects of error propagations in the analysis due to the application to measurements of poor precision.

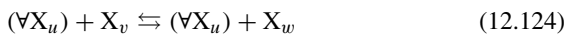
We consider a simple test model which has the advantage that it can be studied analytically even for very large numbers of intermediates, which makes it suitable for the analysis of the interference between experimental errors with the errors due to linearization. This type of model, which is somewhat similar to Eigen's hypercycle model [26], has recently been introduced in connection with a population genetic problem [12]. The model used here is essentially a space-independent, homogeneous version of the model from [12]. We assume that there are two types of chemical species in the system, stable chemicals,  $A_v$ ,  $v = 1, 2, \dots$ , and active intermediates  $X_u$ ,  $u = 1, 2, \dots$ , and that there is a very large supply of stable species  $A_v$ ,  $v = 1, 2, \dots$ , and their concentrations  $a_v$ ,  $v = 1, 2, \dots$ , are assumed to be constant and only the concentrations  $x_u$ ,  $u = 1, 2, \dots$ , of the active intermediates are variable. We consider that the active intermediates replicate, transform into each other, and disappear through autocatalytic processes; moreover we assume that all active intermediates have the same

catalytic activity. Under these circumstances we can represent the chemical processes occurring in the system by the following reaction mechanism:

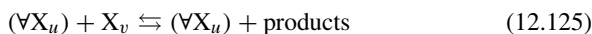
1. Replication processes:



2. Transformation processes:



3. Disappearance processes:



In eqs. (12.123)–(12.125),  $(\forall X_u)$  denotes the pool of active intermediates. The kinetic evolution equations attached to the reactions (12.123)–(12.125) are

$$d\mathbf{x}/dt = x_\Sigma \mathbf{A} \mathbf{x} \quad (12.126)$$

where  $\mathbf{x}$  is the composition vector of the system,  $\mathbf{A}$  is a kinetic matrix that can be expressed in terms of the rate coefficients of the process, and  $x_\Sigma = \sum_u x_u$  is the total concentration of active intermediates. Although the number of variables for this type of systems can be very large, the integration of kinetic equations (12.126) can be reduced to two numerical quadratures. This is made possible by a nonlinear transformation of state variables involving an intrinsic time scale,  $\vartheta(t) = \sum_u \int_0^t x_u(t') dt'$ ; with this intrinsic time scale the kinetic equations (12.126) can be expressed in terms of a linear matrix differential equation:

$$d\mathbf{x}/d\vartheta = \mathbf{A} \mathbf{x} \quad (12.127)$$

If the kinetic matrix is constant, eq. (12.127) can be solved either analytically by using the Sylvester theorem of the Laplace transformation, or numerically. Finally, the relation between the intrinsic time scale  $\vartheta$  and the laboratory time scale  $t$  can be determined from  $t = \int_0^\vartheta [\sum_u x_u(\vartheta')]^{-1} d\vartheta'$ . A constant kinetic matrix is sufficient for representing response experiments for which the excitations can be represented by step functions. For more complicated excitations the kinetic matrix is time-dependent. In this case the method of intrinsic time can be extended by representing the general solution of eq. (12.127) in terms of a Green matrix function  $\mathbf{G}$  which can be evaluated by repeated numerical integration of an equation of the type  $d\mathbf{G}/dt = \mathbf{A}\mathbf{G}$  with initial conditions displaced at different initial times.

As an illustration, we start by analyzing a very simple type of response experiment involving the evaluation of self-replication constants. We consider the perturbation of the self-replication of a species by an input flux, whereas the response to this perturbation is given by the variation of the disappearance rate of the species due to the input perturbation. We study two types of excitation: (a) an increase of the input flux according to a step function and (b) an excitation of the neutral type, where the total input flux is kept constant but the fraction of a labeled compound in the input flux is varied, also as a step function. We assume that the measured values of the variation of the output flux (disappearance rate) are subject to experimental errors. In order to emulate the two types of response experiments, we solve the evolution equations (12.126) exactly for the two

types of response and add exactly the same errors to the two results. These solutions we take as our “experimental” values. Further on, we take the responses with errors and evaluate the self-replication rate from them. For the first type of emulated response experiment, type (a), we evaluate the rate coefficient from the linearized evolution equation. For the neutral response experiment, type (b), we evaluate the self-replication constant by deconvoluting the susceptibility function from our exact response law (12.105). In both cases we study the variation of the relative (percent) output error of the evaluated self-replication rate in terms of the relative (percent) values of the measurement errors and of the relative, percent variations of the input flux. For the experiments of type (a) the variation of the input flux is expressed as the ratio between the flux increase due to excitation and the total value of the flux after the excitation has occurred. For the experiments of type (b) the total flux is kept constant and the variation is the fraction of labeled compound in the input flux after the excitation has occurred.

Figures 12.2 and 12.3 show the error of the evaluated self-replication rate constant as a function of the experimental error and of the variation of the input flux for emulated experiments of type (a) and (b), respectively. In the case of emulated experiments of type (a) for which the evaluation of the rate constant is based on linearized kinetic equations, the error of the evaluated rate constant depends strongly on the variation of the input perturbation. The range of the final output error ( $-40\%$ ,  $+10\%$ ) is distorted in comparison with the range of the experimental error ( $-20\%$ ,  $+20\%$ ). For small values of the input perturbation, between  $20\%$  and  $40\%$ , the output error is surprisingly small—between  $10\%$  and  $0\%$ . As the input perturbation increases, the accuracy of the method deteriorates rapidly and for large perturbations the output error is almost twice as big as the experimental error. For the emulated experiments of type (b), where the rate coefficient is evaluated from our exact response law (12.105) without linearization, the situation is different. For input perturbations between  $20\%$  and  $70\%$  the error of the evaluated rate coefficient has about the same range of variation as the experimental error ( $-20\%$ ,  $+20\%$ ) and does not depend much on the size of the perturbation. For very large input perturbations, between  $70\%$  and  $80\%$ , the output error increases abruptly. In fig. 12.4 we show the difference of errors of the evaluated self-replication

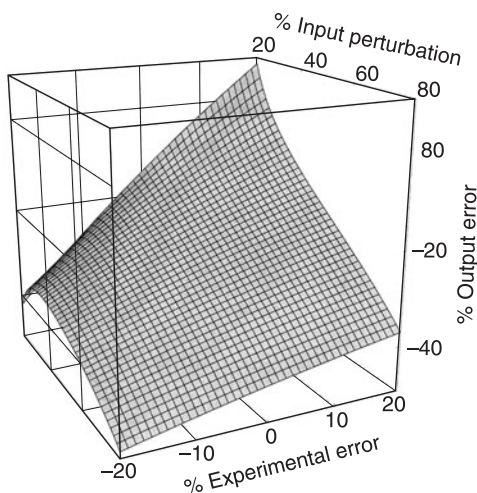


Fig. 12.2 The error of the evaluated self-replication rate constant (output error) versus the experimental error and the input perturbation for an emulated response experiment of type (a). The rate constant is evaluated by using a linearized evolution equation. The range of the output error ( $-40\%$ ,  $+10\%$ ) is strongly distorted in comparison with the range of the experimental error ( $-20\%$ ,  $+20\%$ ). The output error varies greatly with the perturbation size: for large perturbations it is about twice the experimental error, whereas in other regions there is error compensation. (From [11].)

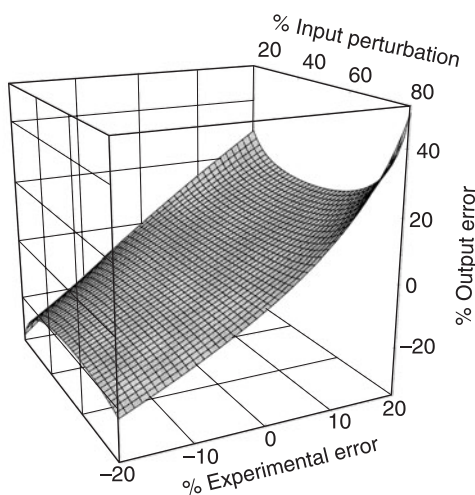


Fig. 12.3 The error of the evaluated self-replication rate constant (output error) versus the experimental error and the input perturbation for an emulated response experiment of type (b). The rate constant is evaluated from the susceptibility function computed from the exact response law. The range of the output error is about the same as the range of the input error for input perturbations between 20% and 70% and does not depend much on the size of the perturbation. For input perturbations between 70% and 80% the output error is bigger. (From [11].)

rate constant (output error) evaluated from emulated experiments of type (a), with linearization, and type (b), without linearization, respectively. We notice that the two approaches lead to different results: the evaluated values of the rate coefficient are different even for small experimental errors and input perturbations. The biggest differences occur for large perturbations, because for large perturbations the linear approach is very inaccurate.

The physical interpretation of the results presented in figs 12.2–12.4 is the following. Like any autocatalytic processes, chemical reactions (12.123)–(12.125) lead to saturation effects due to the balance between self-replication and consumption (disappearance) processes. The saturation effects are nonlinear and as a result the experimental errors propagate nonlinearly, which explains the error distortion displayed in fig. 12.2.

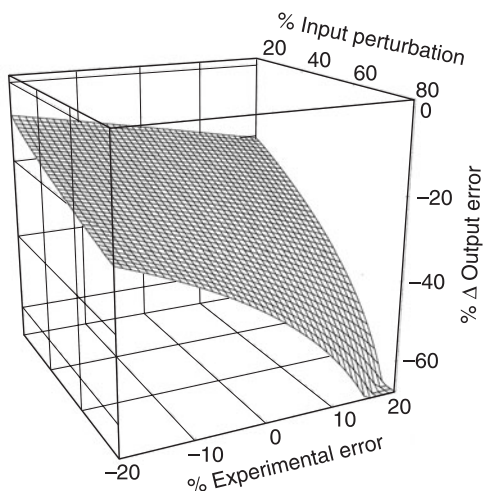


Fig. 12.4 The difference of errors of the evaluated self-replication rate constant (output error) evaluated from emulated experiments of type (a), with linearization, and type (b), without linearization, respectively. The figure shows that the variations of the input perturbation and of the experimental error have a different effect on the two types of response methods. The biggest difference occurs for large perturbations, because for large perturbations the linear approach is very inaccurate. (From [11].)

Even though the analysis of the process is based on linear equations, the intrinsic dynamics of the process is nonlinear and the errors are distorted. This error distortion interferes with the numerical errors due to linearization, resulting in fig. 12.2. In some regions these two errors can have opposite signs, resulting in error compensations, whereas in other regions the two errors add up, leading to large output errors. For neutral response experiments of type (b), linearization is not necessary for the evaluation of the rate coefficients and the most important source of errors are the experimental errors. This is the reason why the output error follows closely the input error and there is almost no error compensation or error amplification.

Numerical errors do, however, have some influence in the case of neutral experiments, especially for very large input perturbations. If the input flux contains a very large fraction of marked (labeled) compound, the dynamics of the marker is close to saturation and in this area small variations of the input flux lead to relatively large variations in the output, which produces numerical errors in the deconvolution of eq. (12.105). This effect explains the spike in fig. 12.3 for large perturbations. It follows that, even for response experiments of type (b), it is not recommended to use very large input perturbations. Nevertheless, it seems that the admissible input perturbations that produce reasonable results are much larger than the admissible perturbations for experiments of type (a) for which the analysis is based on linearized evolution equations.

We have also emulated experiments that involve the measurement of more than one flux, in order to determine the cumulative effect of errors produced by more than one measurement. We have considered the evaluation of two transformation constants. Unfortunately in this case the results cannot be easily represented as compact, 3D graphs. The analysis revealed a complication typical for multiple flux measurements: experimental errors may lead to an apparent violation of the law of mass conservation. We considered only errors that preserve the mass conservation. We noticed the same qualitative features as in the case of the self-replication constant. For experiments of type (a), with linearization, in general there is a distortion of the range of the output error compared to the range of the input error. We have also noticed error compensation and error accumulation; in some cases they exist in more than one region and alternate, one compensation region followed by an accumulation region. Sometimes the amplification error tends to be very large; we have noticed output ranges of the error 4–5 times larger than the maximum range of the measurement errors of the fluxes. The emulated experiments of type (b) are also affected by the cumulative measurement errors in a complicated way. There is a region where the range of the output error is about the same as the maximum range of the input experimental error. However, the maximum admissible input fluxes are smaller than in the case of a single response experiment; in our simulations it varied from approximately 60–65% (compared to 80% for single response experiments) to 35–40%.

Although the detailed results of our computations are probably model specific, we believe that the qualitative features discussed above have general significance. It is likely that the analysis of a response experiment based on linearized equations will lead to error compensation in some regions and to error amplification in other regions. The nonlinear saturation effects occurring in our model are likely to exist even though the autocatalytic steps are missing. For example, enzymatic reactions display saturation effects due to the limited supplies of the enzymes. The error compensation is as harmful as error amplification because in general it is impossible to predict a priori where it occurs.

Concerning the neutral response experiments for which the linearization results are not necessary, it is likely that they lead to better results than the experiments of type (a) for which the analysis is based on the linearization of the kinetic equations. However, even for neutral experiments the use of very large input perturbations requires caution, owing to the proximity to saturation, which results in numerical errors. Multiple measurements with large errors introduce further complications which reduce the potential value of experiments of type (b), without linearization, for large perturbations.

In conclusion, the linearization of the evolution equations for the analysis of chemical and biochemical networks is unpredictably limited. For the linearization to be valid it is necessary to use small perturbations, for which the experimental errors are very large. In the papers that have appeared on this subject, insufficient (or no) attention has been given to error accumulation and propagation [25]. The response approaches developed here avoid linearization and hence are to be preferred.

## 12.6 Response Experiments for Reaction–Diffusion Systems

Our response approach can be extended to inhomogeneous, space-dependent systems. Such a generalization is of interest not only for biochemistry and molecular biology, but also for population biology. In the following we consider a generalized approach, which can be applied to various problems in physics, chemistry, and biology. For simplicity we assume that there is only one fragment in each carrier. We consider a complex system made up of different types of individuals (species) which can be atoms, molecules, quasiparticles, biological organisms, etc. The different types of individuals interact with each other and at the same time are involved in motion, which can be described by transport operators local in time.

We denote by  $\rho_u(\mathbf{r}; t)$ ,  $u = 1, 2, \dots$ , the concentrations of the different species at position  $\mathbf{r}$  and time  $t$ , expressed in numbers of individuals per unit volume, and assume that the rate of change of the species  $u$ ,  $R_u(t)$ , can be expressed as a local, nonlinear, function of the composition vector  $\rho(\mathbf{r}; t) = [\rho_u(\mathbf{r}; t)]_{u=1,2,\dots}$ , and of time  $t$ :

$$R_u(t) = R_u(\rho(\mathbf{r}; t)) = R_u^+(\rho(\mathbf{r}; t)) - R_u^-(\rho(\mathbf{r}; t)), \quad u = 1, 2, \dots,$$

where  $R_u^+(\rho(\mathbf{r}; t)) \geq 0$  and  $R_u^-(\rho(\mathbf{r}; t)) \geq 0$  are formation and consumption rates, respectively. The transport of the different species can be described by transport operators  $\mathbb{L}_u \dots$ , which are local in time and generally nonlocal in space. In this chapter we limit ourselves to transport operators of the “master” type:

$$\mathbb{L}_u \dots = \int_{\mathbf{r}'} [\dots W_u(\rho(\mathbf{r}'; t), \mathbf{r}' \rightarrow \mathbf{r}) d\mathbf{r}' - \dots W_u(\rho(\mathbf{r}; t), \mathbf{r} \rightarrow \mathbf{r}') d\mathbf{r}'] \quad (12.128)$$

where  $W_u(\rho(\mathbf{r}'; t), \mathbf{r}' \rightarrow \mathbf{r})$  are concentration-dependent transition rates. The operators describe the regular (Fick) as well as anticrowding diffusion processes as particular cases. The evolution equations of the process are:

$$\partial \rho_u(\mathbf{r}, t) / \partial t = J_u^+(\mathbf{r}, t) - J_u^-(\rho(\mathbf{r}, t)) + R_u[\rho(\mathbf{r}, t)] + \mathbb{L}_u \rho_u(\mathbf{r}, t), \quad (12.129)$$

where  $J_u^+(\mathbf{r}, t)$  is the input flux of species  $u$ , which is generally space- and time-dependent and can be controlled by the experimenter, and  $J_u^-(\rho(\mathbf{r}, t))$  is the output



flux of species  $u$  which is assumed to depend on the composition vector. Each species  $u = 1, 2, \dots$  may exist in two different forms, “marked” and “not marked,” and both forms fulfill a “neutrality condition” similar to the one introduced for homogeneous systems, that is, their kinetic and transport properties are identical.

In chemical kinetics a “marked species” can be a molecule containing a radioactive isotope and we neglect the kinetic isotope effect. In fluid mechanics a “marked species” can be a colored fluid for which the hydrodynamic properties (density, viscosity, diffusion coefficients) are the same as the ones of the main fluid. In population genetics a “marked species” can be an individual carrying a neutral mutation, and for which the main functions describing the vital statistics (natality and mortality functions, diffusion coefficients) are the same as in the case of a nonmutant individual. In the following we denote by  $\rho_u(\mathbf{r}, t)$  and  $\rho_u^*(\mathbf{r}, t)$ ,  $u = 1, 2, \dots$ , the concentrations of the “not marked” and “marked” species, respectively, and by  $\rho_u^\Sigma(\mathbf{r}, t) = \rho_u(\mathbf{r}, t) + \rho_u^*(\mathbf{r}, t)$ ,  $u = 1, 2, \dots$ , the total concentrations of the species.

At the beginning of the experiment the system contains only “not marked” species. The system need not be but may be in a stationary state. The experiment consists in varying the ratios

$$\alpha_u(\mathbf{r}, t) = J_u^{+*}(\mathbf{r}, t) / J_u^{+\Sigma}(\mathbf{r}, t) \quad (12.130)$$

between the input fluxes of  $J_u^{+*}(\mathbf{r}, t)$ , the “marked” compounds, and the total input fluxes,  $J_u^{+\Sigma}(\mathbf{r}, t)$ , with the preservation of the total input fluxes  $J_u^{+\Sigma}(\mathbf{r}, t)$ . We record the response to these variations, the fractions

$$\beta_u(\mathbf{r}, t) = J_u^{-*}[\rho(\mathbf{r}, t), \rho^*(\mathbf{r}, t)] / J_u^{-\Sigma}[\rho(\mathbf{r}, t), \rho^*(\mathbf{r}, t)] \quad (12.131)$$

of the marked outflow fluxes,  $J_u^{-*}[\rho(\mathbf{r}, t), \rho^*(\mathbf{r}, t)]$ , to the total output fluxes,  $J_u^{-\Sigma}[\rho(\mathbf{r}, t), \rho^*(\mathbf{r}, t)]$ . We intend to obtain a relation between the excitation of the system, expressed by the fractions  $\alpha_u(\mathbf{r}, t)$ , and the response of the system, expressed by the fractions  $\beta_u(\mathbf{r}, t)$ .

We assume the existence of generalized “neutrality conditions” [8,11], in the form of scaling laws, which connect the kinetic and transport laws for the whole system to the corresponding laws for the “marked” and “not marked” species, respectively:

$$R_u^{\pm*}(\rho^*(\mathbf{r}; t), \rho(\mathbf{r}; t)) = \frac{\rho_u^*(\mathbf{r}; t)}{\rho_u^*(\mathbf{r}; t) + \rho_u(\mathbf{r}; t)} R_u^{\pm}(\rho^*(\mathbf{r}; t) + \rho(\mathbf{r}; t)), \quad u = 1, 2, \dots \quad (12.132)$$

Equations (12.132) express the fact that the “marked” and “not marked” species contribute equally to the transport process. We assume that the output fluxes,  $J_u^{-\Sigma}[\rho(\mathbf{r}, t)]$ , are expressed by kinetic laws, and thus they also obey a scaling condition similar to eq. (12.132):

$$J_u^{-*}(\rho^*(\mathbf{r}; t), \rho(\mathbf{r}; t)) = \frac{\rho_u^*(\mathbf{r}; t)}{\rho_u^*(\mathbf{r}; t) + \rho_u(\mathbf{r}; t)} J_u^{-\Sigma}(\rho^*(\mathbf{r}; t) + \rho(\mathbf{r}; t)) \quad (12.133)$$

We also introduce similar scaling conditions for the transport rates  $W_u(\rho(\mathbf{r}'; t), \mathbf{r}' \rightarrow \mathbf{r}) d\mathbf{r}$ ,  $u = 1, 2, \dots$ :

$$W_u^*(\rho^*(\mathbf{r}; t), \rho(\mathbf{r}'; t), \mathbf{r}' \rightarrow \mathbf{r}) d\mathbf{r} = \frac{\rho_u^*(\mathbf{r}'; t)}{\rho_u^*(\mathbf{r}'; t) + \rho_u(\mathbf{r}'; t)} \times W_u^*(\rho^*(\mathbf{r}'; t) + \rho(\mathbf{r}'; t), \mathbf{r}' \rightarrow \mathbf{r}) d\mathbf{r}, \quad u = 1, 2, \dots \quad (12.134)$$

The response experiment can be described in terms of two sets of nonlinear evolution equations of the type (12.129), one set for the labeled concentration vector  $\rho^*(\mathbf{r}; t)$ , and the second set for the total concentration vector  $\rho(\mathbf{r}; t) + \rho^*(\mathbf{r}; t)$ , respectively. The transport and reaction rates in these two sets of evolution equations are connected to each other by means of the neutrality conditions (12.132)–(12.134). Together with suitable initial and boundary conditions, these two sets of evolution determine the time and space dependence of the total concentrations of the different species, as well as the concentrations of the marked species. Despite their nonlinearity, these two sets of evolution equations lead to a linear response law, which relates the excitation functions  $\alpha_u(\mathbf{r}, t)$  to the response functions  $\beta_u(\mathbf{r}, t)$ . By using the neutrality conditions (12.132)–(12.134) and the definitions (12.130)–(12.131) for the excitation and response functions, after lengthy algebraic manipulations, the two sets of nonlinear evolution equations lead to a set of linear integro-differential equations, which relate the responses  $\beta_u(\mathbf{r}, t)$  to the excitations  $\alpha_u(\mathbf{r}, t)$ . By representing the solution of this set of linear equations in terms of Green functions, we come to a generalization of eq. (12.105):

$$\beta_u(\mathbf{r}, t) = \sum_{u'} \int_{-\infty}^t \int_{\mathbf{r}'} \chi_{uu'}(\mathbf{r}', t' \rightarrow \mathbf{r}, t) \alpha_{u'}(\mathbf{r}', t') d\mathbf{r}' dt' \quad (12.135)$$

where the space- and time-dependent susceptibility function  $\chi_{uu'}(\mathbf{r}', t' \rightarrow \mathbf{r}, t)$  is nonnegative and obeys the normalization condition:

$$\sum_{u'} \int_{-\infty}^t \int_{\mathbf{r}'} \chi_{uu'}(\mathbf{r}', t' \rightarrow \mathbf{r}, t) d\mathbf{r}' dt' = 1 \quad (12.136)$$

It is possible to show that the susceptibility function has a physical interpretation, similar to the one existing for homogeneous systems. It is related to the probability density

$$\varphi_{u'}(\tau, \Delta\mathbf{r}|\mathbf{r}, u; t) d\tau d\Delta\mathbf{r}, \quad \text{with} \quad \sum_{u'} \int_0^\infty \int_{\mathbf{r}'} \varphi_{u'}(\tau, \Delta\mathbf{r}|\mathbf{r}, u; t) d\tau d\Delta\mathbf{r} = 1 \quad (12.137)$$

that a species  $u$  which leaves the system at time  $t$  and position  $\mathbf{r}$  entered the system as the species  $u'$ , spent in the system a residence time between  $\tau$  and  $\tau + d\tau$ , with  $\tau = t - t'$ , and has a displacement vector between  $\Delta\mathbf{r}$  and  $\Delta\mathbf{r} + d\Delta\mathbf{r}$ , with  $\Delta\mathbf{r} = \mathbf{r} - \mathbf{r}'$ . We have

$$\varphi_{u'}(\tau, \Delta\mathbf{r}|\mathbf{r}, u; t) = \chi_{uu'}(\mathbf{r} - \Delta\mathbf{r}, t - \tau \rightarrow \mathbf{r}, t) \quad (12.138)$$

According to eq. (12.138) the physical significance of the linear response law (12.135) is straightforward: it expresses the contribution to the output fraction of marked species entering the system in different forms  $u'$ , at different initial positions  $\mathbf{r}'$  and different initial times,  $t' = t - \tau$ . The weight function (susceptibility function)

attached to various initial positions and times is the conditional probability density of these three random variables. The theoretical results presented here include as particular cases our previous approaches to linear response for nonlinear systems with neutrality conditions [1–12]. In order to save space the derivation of eqs. (12.135)–(12.138) is not given here. Similar derivations, corresponding to some particular cases, are given in our previous publications [1–12].

For illustrating our approach to response in inhomogeneous systems, we consider the problem of enhanced (hydrodynamic) transport induced by population growth in reaction–diffusion systems and its application to human population genetics. The study of geographical distributions of various mutations provides useful clues to human evolution. It is interesting that many mutations are neutral and that the spreading of these mutations in the total population can be interpreted as a naturally generated neutral response experiment.

The spreading of a mutation in a migrating population may display enhanced (hydrodynamic) transport induced by population growth, a phenomenon that can occur not only in population genetics but also in physics and chemistry. We start by presenting the problem of enhanced transport on reaction–diffusion systems and then proceed with the study of propagation of neutral mutations in human populations. We consider a system made up of different individuals  $X_u$ ,  $u = 1, 2, \dots$  (molecules, quasiparticles, biological organisms, etc.). The species  $X_u$ ,  $u = 1, 2, \dots$ , replicate, transform into each other, die, and at the same time undergo slow, diffusive motion, characterized by the diffusion coefficients  $D_u$ ,  $u = 1, 2, \dots$ , which are assumed to be constant. The replication and disappearance rates  $R_u^\pm$  of the different species are assumed to be proportional to the species densities  $X_u$ ,  $u = 1, 2, \dots$ , we have  $R_u^\pm = x_u \rho_u^\pm(\mathbf{x})$ , where the rate coefficients  $\rho_u^\pm(\mathbf{x})$  are generally dependent on the composition vector  $\mathbf{x} = (x_u)$ ; similarly, the rate  $R_{u \rightarrow v}$  of transformation of species  $X_u$  into the species  $X_v$  is given by  $R_{u \rightarrow v} = x_u k_{uv}(\mathbf{x})$ , where  $k_{uv}(\mathbf{x})$  are composition-dependent rate coefficients. Under these circumstances the process can be described by the following reaction–diffusion equations:

$$\frac{\partial}{\partial t} x_u = x_u \rho_u^+(\mathbf{x}) - x_u \rho_u^-(\mathbf{x}) + \sum_{v \neq u} [x_v k_{vu}(\mathbf{x}) - x_u k_{uv}(\mathbf{x})] + D_u \nabla^2 x_u \quad (12.139)$$

We are interested in the time and space evolution of the fractions of the different species present in the system:  $\gamma_u = x_u/x$ , with  $1 = \sum_u \gamma_u$ , where  $x = \sum_u x_u$  is the total population density. For example, in chemistry  $\gamma_u$  are molar fractions whereas in population genetics they are gene frequencies. After lengthy algebraic transformations, eqs. (12.139) lead to the following evolution equations for the total population density  $x$  and for the fractions  $\gamma_u$ :

$$\frac{\partial}{\partial t} x = x [\tilde{\rho}^+(x, \gamma) - \tilde{\rho}^-(x, \gamma)] + \nabla^2 (x \tilde{D}(\gamma)) \quad (12.140)$$

$$\frac{\partial}{\partial t} \gamma_u + \nabla (\mathbf{v}_u \gamma_u) = D_u \nabla^2 \gamma_u + \varepsilon_u \gamma_u + \sum_{v \neq u} [\gamma_v k_{vu}(x\gamma) - \gamma_u k_{uv}(x\gamma)] + \delta \mathcal{R}_u \quad (12.141)$$

where

$$\tilde{\rho}^{\pm}(x, \gamma) = \sum_u \gamma_u \rho_u^{\pm}(x\gamma), \quad \tilde{D}(\gamma) = \sum_u \gamma_u D_u \quad (12.142)$$

are average rate and transport coefficients,

$$\delta\rho_u^{\pm}(x, \gamma) = \rho_u^{\pm}(x\gamma) - \tilde{\rho}^{\pm}(x, \gamma), \quad \delta D_u(\gamma) = D_u - \tilde{D}(\gamma) \quad (12.143)$$

are deviations of the individual rate and transport coefficients from the corresponding average values,

$$\mathbf{v}_u = -2D_u \nabla \ln x, \quad \varepsilon_u = \text{div}(\mathbf{v}_u) \quad (12.144)$$

are transport (hydrodynamic) speeds and expansion coefficients attached to different population fractions,

$$\begin{aligned} \delta\mathcal{R}_u = & \gamma_u [\delta\rho_u^{+}(x, \gamma) - \delta\rho_u^{-}(x, \gamma)] - \gamma_u \left\{ \sum_v \delta D_v(\gamma) [\nabla^2 \gamma_v + 2(\nabla \ln x) \bullet \nabla \gamma_v] \right\} \\ & + \delta D_u(\gamma) \gamma_u x^{-1} \nabla^2 x \end{aligned} \quad (12.145)$$

are the components of the rates of change of the population fractions due to the individual variations of the rate and transport coefficients, and  $\gamma = (\gamma_1, \gamma_2, \dots)$  is the vector of population fractions.

We notice that, even though the different species are undergoing slow, diffusive motions, the corresponding population fractions move faster: in the evolution equations (12.141) there are both diffusive terms as well as convective transport (hydrodynamic) terms depending on the transport speeds  $\mathbf{v}_u$  given by eqs. (12.145). According to eqs. (12.145) these transport speeds are generated by the space variations of the total population density and have the opposite sign of the gradient of the total population densities. For a growing population the population cloud usually expands from an original area and tries to occupy all space available. The population density decreases toward the edge of the population cloud; thus the population gradient is negative and the transport velocities are positive, oriented toward the directions of propagation of the population cloud. It follows that the cause of enhanced transport of the species fractions is the net population growth. Since the gradient tends to increase toward the edge of the population wave, an initial perturbation of the species fractions generated in the propagation front of the population has a good chance of undergoing enhanced transport and spreading all over the system. An initial perturbation produced close to the initial area where the population originates has a poor chance of undergoing sustained enhanced transport.

It is important to clarify the mathematical and physical significance of the hydrodynamic transport terms  $\nabla(\mathbf{v}_u \gamma_u)$  in eqs. (12.141). From the mathematical point of view the terms  $\nabla(\mathbf{v}_u \gamma_u)$  emerge as a result of a nonlinear transformation of the state variables, from species densities to species fractions. The physical interpretation of the transport terms  $\nabla(\mathbf{v}_u \gamma_u)$  depends on the direction and orientation of the speed vectors: for expanding populations  $\mathbf{v}_u$  are generally oriented toward the direction of expansion of the population cloud, resulting in enhanced transport. For shrinking population clouds the terms  $\nabla(\mathbf{v}_u \gamma_u)$  lead to the opposite effect, that of the transport process slowing down.

In general, different population fractions have different propagation speeds. An interesting particular case is that for which the replication and disappearance rate coefficients

and the diffusion coefficients are the same for all species and depend only on the total population density  $\rho_u^\pm(\mathbf{x}) = \rho^\pm(x)$ ,  $D_u = D$ . Moreover, we assume that the transformation rates are constant,  $k_{uv}(\mathbf{x}) = k_{uv}$ . This type of condition is fulfilled in chemistry by tracer experiments, for which the variation of the rate and transport coefficients due to the kinetic isotope effect can be neglected [18]. Similar restrictions are fulfilled in population genetics, in the case of neutral mutations, for which the demographic and transport parameters are the same for neutral mutants and nonmutants, respectively [12]. For neutral systems the evolution equations turn into a simpler form:

$$\frac{\partial}{\partial t}x = x\mu(x) + D\nabla^2x \quad (12.146)$$

$$\frac{\partial}{\partial t}\gamma_u + \nabla(\mathbf{v}\gamma_u) = D\nabla^2\gamma_u + \varepsilon\gamma_u + \sum_{v \neq u}(\gamma_vk_{vu} - \gamma_uk_{uv}) \quad (12.147)$$

where  $\mu(x) = \rho^+(x) - \rho^-(x)$  is the net production rate of the total population. We notice that the total population density obeys a separate equation, which is independent of the species fractions, and the evolution equations for the fractions become linear.

Now we can investigate the problem which suggested the present research, the geographical spreading of neutral mutations in human populations [12,27]. We consider a growing population which diffuses slowly in time and assume that the net rate of growth is a linear function of population density,  $\mu(x) = \mathfrak{r}_L(1 - x/x_\infty)$ , where  $\mathfrak{r}_L$  is Lotka's intrinsic rate of growth of the population. We assume that, at an initial position and time, a neutral mutation occurs and afterwards no further mutations occur. We are interested in the time and space dependence of the local fractions of the individuals, which are the offspring of the individual that carried the initial mutation. The ultimate goal of this analysis is the evaluation of the position and time where the mutation originated from measured data representing the current geographical distribution of the mutation. We limit our analysis to one-dimensional systems, for which a detailed theoretical analysis is possible. Equations (12.146)–(12.147) become

$$\frac{\partial}{\partial t}x = \mathfrak{r}_Lx \left(1 - \frac{x}{x_\infty}\right) + D\nabla^2x \quad (12.148)$$

$$\frac{\partial}{\partial t}\gamma + \nabla(\mathbf{v}\gamma) = D\nabla^2\gamma + \varepsilon\gamma \quad (12.149)$$

where  $\gamma$  is the local fraction of mutants. Equation (12.148) for the total population growth is the well-known Fisher equation [28] which has solitary solutions.

Two different solitary wave solutions for eq. (12.148) have been derived in the literature. There is an asymptotic solution developed by Fisher and others [28] and an exact solution derived by Ablowitz and Zeppetella [29]. The asymptotic solution gives an excellent representation of the population wave from the top saturation level to the front edge where the total population tends to zero. The usefulness of the exact solution of Ablowitz and Zeppetella has been questioned in the literature [28] because, although exact, it does not represent all possible solutions and the most relevant solutions may not be represented by it. In the following we consider only the Fisher solution, which is biologically significant [12]. For the application of the Ablowitz and Zeppetella solution,

see [12]. In our notation the asymptotic solution is

$$x(r, t) = x_\infty \left\{ 1 + \exp \left[ \frac{c}{2D} (r - ct) \right] \right\}^{-1}, \quad \text{with } c = 2\sqrt{\hbar_L D} \quad (12.150)$$

The transport speed  $v$  and for the expansion factor  $\varepsilon$  are given by

$$\frac{v}{c} = 1 - \frac{x}{x_\infty} = \frac{2 \exp \left[ \frac{c}{4D} (r - ct) \right]}{\cosh \left[ \frac{c}{4D} (r - ct) \right]} \quad (12.151)$$

$$\varepsilon = \frac{c^2}{4D} \left\{ \cosh \left[ \frac{c}{4D} (r - ct) \right] \right\}^{-2} \geq 0 \quad (12.152)$$

For a mutation that occurs at the very edge of the total population wave,  $x \sim 0$ , the transport speed is equal to the speed of propagation front,  $v = c$ ; the motion of two waves, for the total population and the mutation, is synchronized.

Now we can address the problem of determining the position where a mutation originates. The probability density  $P(\mathbf{r}, t)$  of the position of the center of gravity of the mutant population can be roughly estimated by normalizing the mutant gene frequency:

$$P(r, t) \sim \gamma(r, t) / \int \gamma(r, t) dr \quad (12.153)$$

By applying eq. (12.153) we obtain a Gaussian distribution for the position  $r$  of the center of gravity of the mutant population both for the diffusive regime and for the enhanced transport:

$$P(r, t) \sim [4\pi D (t - t_0)]^{-1/2} \exp \left\{ - (r - r_0 - v_{\text{eff}} (t - t_0))^2 / [4D (t - t_0)] \right\} \quad (12.154)$$

where  $r_0$  and  $t_0$  are the initial position and time where the mutation occurred and  $v_{\text{eff}}$  is an effective transport rate. In eq. (12.154) we have  $v_{\text{eff}} \sim 0$  for diffusive transport,  $x \sim x_\infty$ , and  $v_{\text{eff}} \sim c$  for the enhanced transport,  $x \sim 0$ . Although for intermediate cases between these two extremes the probability density  $P(r, t)$  is generally not Gaussian, eq. (12.154) can be used as a reasonable approximation, where the effective propagation speed  $v_{\text{eff}}$  has an intermediate value between zero (diffusive transport) and the maximum values for enhanced transport corresponding to the two solutions of the Fisher equation. Under these circumstances the average position of the center of gravity of the mutant population increases linearly in time. From eq. (12.154) we obtain:

$$\langle r(t) \rangle = \int r P(r, t) dr = r_0 + v_{\text{eff}} (t - t_0) \quad (12.155)$$

If we examine the current geographical distribution of a mutation it is hard to estimate the value of the population density  $x$  at the position and time where the mutation originates. Under these circumstances it makes sense to treat  $x$  as a random variable selected from a certain probability density  $p(x)$ . The only constraints imposed on  $p(x)$  are the conservation of the normalization condition  $\int p(x) dx = 1$ , and the range of variation,  $x_\infty \geq x \geq 0$ . By using the maximum information entropy approach, we can show that the most unbiased probability density  $p(x)$  is the uniform one:

$$p(x) = (x_\infty)^{-1} [h(x) - h(x - x_\infty)] \quad (12.156)$$

where  $h(x)$  is the Heaviside step function. By considering a large sample of different initial conditions,  $v_{\text{eff}}$  can be evaluated as an average value:

$$v_{\text{eff}} = \int v(x) p(x) dx \quad (12.157)$$

where  $v(x)$  is given by eqs. (12.151). We have

$$v_{\text{eff}} = c \int_0^{x_\infty} \left(1 - \frac{x}{x_\infty}\right) \frac{dx}{x_\infty} = \frac{1}{2}c \quad (12.158)$$

For the estimation of the initial position of a mutation it is useful to consider the ratio

$$\zeta = \frac{r(t_L) - r(t_0)}{\langle r(t_L) \rangle - r(t_0)} \quad (12.159)$$

where  $r(t_L)$  is the position of the limit of expansion,  $t_L$  is the time necessary for reaching the limit of expansion,  $\langle r(t_L) \rangle$  is the position of the center of gravity of the mutant population for  $t = t_L$ , and  $r(t_0) = r_0$  is the position where the mutation originates. In eq. (12.159) both  $r(t_L)$ , the position of the limit of expansion, and  $\langle r(t_L) \rangle$ , the current average position of the center of gravity of the mutant population, are accessible experimentally. It follows that, if the ratio  $\zeta$  can be evaluated from theory, then  $r(t_0) = r_0$ , the point of origin of the mutation, can be evaluated from eq. (12.159). By taking into account that the total population wave moves with the speed  $c$  and that the average center of gravity moves with the speed  $v_{\text{eff}}$ , we have

$$\zeta = \frac{ct_L - ct_0}{r(t_0) + v_{\text{eff}}(t_L - t_0) - r(t_0)} = \frac{c}{v_{\text{eff}}} \quad (12.160)$$

From eqs. (12.157) and (12.160) it follows that  $\zeta = 2$ , a value that is in good agreement with the numerical simulations of Edmonds et al. [27], which lead to  $\zeta = 2.2$ . The difference of 0.2 between theory and simulations is due to the random drift, which was taken into account in the simulations but is neglected in our theory. By including the random drift, our theory provides information about the details of the motion of the propagation front [12].

In conclusion, we have shown that the neutral response approach can be extended to inhomogeneous, space-dependent reaction–diffusion systems. For labeled species (tracers) that have the same kinetic and transport properties as the unlabeled species, there is a linear response law even if the transport and kinetic equations of the process are nonlinear. The susceptibility function in the linear response law is given by the joint probability density of the transit time and of the displacement position vector. For illustration we considered the time and space spreading of neutral mutations in human populations and have shown that it can be viewed as a natural linear response experiment. We have shown that enhanced (hydrodynamic) transport due to population growth may exist and developed a method for evaluating the position of origin of a mutation from experimental data.

## 12.7 Conclusions

In this chapter we suggested new types of experiments for large chemical and biochemical systems which lead to linear response even if the perturbations are large and the underlying kinetics is highly nonlinear. We have shown that the susceptibility functions in the response equations have a simple physical interpretation: they are given by the probability densities of the transit times and displacement vector of a given fragment crossing the system. We have developed methods for extracting connectivity, mechanistic, and kinetic information from response experiments.

Because of space limitations we considered only macroscopic kinetic systems, but the approach can be also applied to experiments of single-molecule kinetics [10]. The suggested approach is in an early stage of development. An important issue is extending the approach to nonneutral systems, for which the kinetic and transport properties of the labeled species are different from the corresponding properties of the unlabeled species. There are two different regimes: (a) If the deviations from neutrality are small, the response can be represented by a functional Taylor expansion; the terms of first order in the functional Taylor expansion correspond to the linear response law. (b) For large deviations, a phase linearization approach is more appropriate.

## References

- [1] Moran, F.; Vlad, M. O.; Ross, J. Transition and transit time distributions for time-dependent reactions with application to biochemical networks. *J. Phys. Chem. B* **1997**, *15*, 296–303.
- [2] Vlad, M. O.; Moran, F.; Ross, J. *H*-Theorem for lifetime distributions of active intermediates in nonequilibrium chemical systems with stable limit cycles. *J. Phys. Chem. B* **1999**, *102*, 4598–4611.
- [3] Vlad, M. O.; Ross, J.; Huber, D. L. Linear free energy relations and reversible stretched exponential kinetics in systems with static and dynamic disorder. *J. Phys. Chem. B* **1999**, *103*, 1563–1580.
- [4] Vlad, M. O.; Moran, F.; Ross, J. Transit time distributions for biochemical networks far from equilibrium: amplification of the probability of net transformation due to multiple reflections. *J. Phys. Chem. B* **1999**, *103*, 3965–3974.
- [5] Vlad, M. O.; Moran, F.; Ross, J. Response theory for random channel kinetics in complex systems: application to lifetime distributions of active intermediates. *Physica A* **2000**, *278*, 504–525.
- [6] Vlad, M. O.; Cerofolini, G. F.; Ross, J. Statistical fractal adsorption isotherms, linear energy relations, and power-law trapping-time distributions in porous media. *Phys. Rev. E* **2000**, *62*, 837–844.
- [7] Vlad, M. O.; Moran, F.; Ross, J. Effective medium approximation of Taylor transport in systems with static disorder. *J. Phys. Chem. B* **2001**, *105*, 11710–11718.
- [8] Vlad, M. O.; Moran, F.; Tsuchiya, M.; Cavalli-Sforza, L. L.; Oefner, P. J.; Ross, J. Neutrality condition and response law for nonlinear reaction-diffusion equations, with application to population genetics. *Phys. Rev. E* **2002**, *65*, 1–17.
- [9] Vlad, M. O.; Moran, F.; Rodriguez, Y.; Ross, J. Delayed response in tracer experiments and fragment-carrier approach to transit time distributions in nonlinear chemical kinetics. *Int. J. Bifurcation & Chaos* **2002**, *12*, 2599–2618.
- [10] Vlad, M. O.; Moran, F.; Ross, J. Lifetimes and on-off distributions for single molecule kinetics: stochastic approach and extraction of information from experimental data. *Chem. Phys. (Elsevier)* **2003**, *287*, 83–90.
- [11] Vlad, M. O.; Arkin, A.; Ross, J. Response experiments for nonlinear systems with application to reaction kinetics and genetics. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7223–7228.



- [12] Vlad, M. O.; Cavalli-Sforza, L. L.; Ross, J. Enhanced (hydrodynamic) transport induced by population growth in reaction–diffusion systems with application to population genetics. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10249–10253.
- [13] Sochet, L. R. *La cinétique des réactions en chaînes*; Dunod: Paris, 1971; p 22.
- [14] Burnett, G. M. *Mechanism of Polymer Reactions*; Interscience: New York, 1954; Chapter 7.
- [15] Iosifescu, M.; Tautu, P.; *Stochastic Processes and Applications in Biology and Medicine*, Vol. II: *Models*; Springer: Berlin, 1973; Chapter 1.
- [16] Morse, P. M.; Feshbach, H. *Methods of Theoretical Physics*, Vol. 1; McGraw-Hill: New York, 1953; Chapter 6.
- [17] Drazer, G.; Zanette, D. Z. Experimental evidence of power-law trapping-time distributions in porous media. *Phys. Rev. E* **1999**, *60*, 5858.
- [18] Neiman, M. B.; Gál, D. *The Kinetic Isotope Method and Its Application*; Akadémia Kiadó: Budapest, 1971.
- [19] Montroll, E. W.; West, B. J. In *Fluctuation Phenomena: Studies in Statistical Mechanics*, Vol. VII; North-Holland: Amsterdam, 1979; pp 61–175 and references therein.
- [20] Rudzinski, W.; Evrett, H. D. *Adsorption of Gases on Heterogeneous Surfaces*; Academic Press: London, 1992.
- [21] Boudart, M. *Kinetics of Chemical Processes*; Prentice-Hall: Englewood Cliffs, NJ, 1968; Kosower, E. M. *An Introduction to Physical Organic Chemistry*; Wiley: New York, 1968; Chapman, N. B.; Shorter, J. *Advances in Linear Free Energy Relationships*; Plenum Press: London, 1972.
- [22] Boudart, M. In *Physical Chemistry: An Advanced Treatise*, Vol. VII: *Reactions in Condensed Phases*; Eyring, H., Ed.; Academic Press: New York, 1975; pp 349–411.
- [23] Easterby, J. S. The analysis of metabolite channelling in multienzyme complexes and multifunctional proteins. *Biochem. J.* **1989**, *264*, 605–607. Easterby, J. S. Integration of temporal analysis and control analysis of metabolic systems. *Biochem. J.* **1990**, *269*, 255–259.
- [24] Aris, R. *Vectors, Tensors and the Basic Equation of Fluid Mechanics*; Dover: New York, 1990.
- [25] Yeung, M. K. S.; Tegner, J.; Collins, J. J. Reverse engineering gene networks using singular value decomposition and robust regression. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6163–6168.
- [26] Eigen, M. *The Hypercycle: A Principle of Natural Self Organization*; Springer: Berlin, 1979.
- [27] Edmonds, C. A.; Lillie, A. S.; Cavalli-Sforza, L. L.; Mutations arising in the wave front of an expanding population. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 975–979.
- [28] Fisher, R. A. The wave of advance of advantageous genes. *Ann. Eugen.* **1937**, *7*, 353–369; Murray, J. D. *Mathematical Biology*; Springer, New York, 3rd ed., Vol. 1, 2002; Chapter 13.
- [29] Ablowitz, M. J.; Zeppetella, A. Explicit solutions of Fisher’s equation for a special wave speed. *Bull. Math. Biol.* **1979**, *41*, 835–840.

## Mini-Introduction to Bioinformatics

---

There is enormous interest in the biology of complex reaction systems, be it in metabolism, signal transduction, gene regulatory networks, protein synthesis, and many others. The field of the interpretation of experiments on such systems by application of the methods of information science, computer science, and biostatistics is called bioinformatics (see [1] for a presentation of this subject). Part of it is an extension of the chemical approaches that we have discussed for obtaining information on the reaction mechanisms of complex chemical systems to complex biological and genetic systems. We present here a very brief introduction to this field, which is exploding with scientific and technical activity. No review is intended, only an indication of several approaches on the subject of our book, with apologies for the omission of vast numbers of publications.

A few reminders: The entire complement of DNA molecules constitute the genome, which consists of many genes. RNA is generated from DNA in a process called transcription; the RNA that codes for proteins is known as messenger RNA, abbreviated to mRNA. Other RNAs code for functional molecules such as transfer RNAs, ribosomal components, and regulatory molecules, or even have enzymatic function. Protein synthesis is regulated by many mechanisms, including that for transcription initiation, RNA splicing (in eukaryotes), mRNA transport, translation initiation, post-translational modifications, and degradation of mRNA. Proteins perform perhaps most cellular functions.

Advances in microarray technology, with the use of cDNA or oligonucleotides immobilized in a predefined organization on a solid phase, have led to measurements of mRNA expression levels on a genome-wide scale (see chapter 3). The results of the measurements can be displayed on a plot (fig. 13.1) on which a row represents one gene at various times, a column the whole set of genes, and the time of gene expression

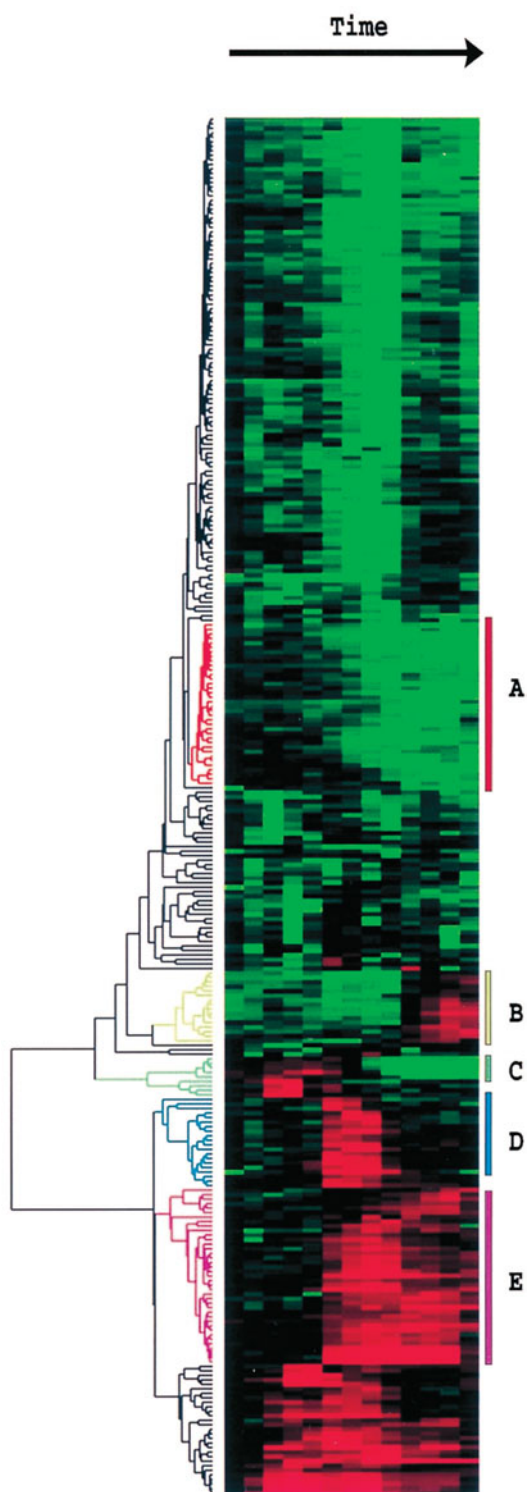


Fig. 13.1 Clustering display of data from time course of expression by stimulation of primary human fibroblasts. See p. 209 for full caption.

is plotted along the axis of rows. The changes in expression levels, as measured by fluorescence, are indicated by colors, for example green for decreased expression, black for no change in expression, and red for increased expression. Responses in expression levels have been measured for various biochemical and physiological conditions. We turn now to a few methods of obtaining information on genomic networks from microarray measurements.

### 13.1 Clustering

Clustering is the grouping of genes into hierarchical functional units based on similarities—correlations—in expression patterns. We used this method in chapter 7 (see figs. 7.6 and 7.9 and the associated discussion). An example of clustering is given in fig. 13.1, taken from [2]. Clustering is a useful method of visualization of large banks of data rather a method of reverse engineering, that is, the deduction of gene networks from measurements. There is the implication in clustering that genes that are expressed in similar patterns are mechanistically related, or perform related biological functions. Both this assumption and the one of hierarchical function may in some cases, but not all, be valid. There is evidence that gene regulatory networks are not hierarchical but are rather interwoven, as is the case for metabolic and protein networks.

### 13.2 Linearization in Various Forms

The kinetic equations governing gene expression can be expected to be highly non-linear and therefore difficult to investigate. Linearization of the kinetic equations is tempting and certainly facilitates matters. A number of investigations [3–5] have studied such an approach and we discuss briefly one of them.

Young et al. [4] propose a reverse engineering method in which they start with general chemical kinetic equations (see eq. (11.1)) and then linearize these equations (see eq. (11.3)) for the concentrations of each mRNA in the system: the time variation of each mRNA is a linear function of its own concentration; a linear function of all other mRNA concentrations with coefficients that give the type (positive, negative, or zero) and strength of interaction between two mRNAs; a constant perturbation term (external stimulus); and a noise term dependent on time only. The coefficients form the connectivity matrix. The system of these  $N$  mRNAs is originally presumed to be in a stationary

---

Fig. 13.1 Clustering display of data from time course of expression by stimulation of primary human fibroblasts. Serum fibroblasts were grown in culture, deprived of serum for 48 hours; serum was then added and samples were taken at time intervals from 0 to 24 hours. Microarray data was taken from the expression of about 8,600 genes. All measurements are relative to time 0. The color scale ranges from saturated green for log ratios  $-3.0$  and below, to saturated red for log ratios  $3.0$  and above. Each gene is represented by a single row, and each time point is represented by a single column. Five clusters are indicated by colored bars. The genes in A are involved in cholesterol biosynthesis, in B in the cell cycle, in C in the immediate-early response, in D in signaling and angiogenesis, and in E in wound healing and tissue remodeling. (The picture and a part of the caption are taken from [2], with permission.) (*See color insert.*)

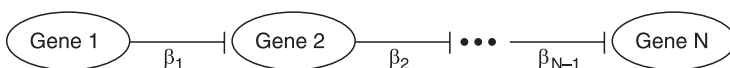


Fig. 13.2 Schematic of a one-dimensional gene network with cascade structure. (Taken from [4], with permission.)

(nonequilibrium) state, from which it is perturbed a small amount so that the responses of all mRNAs, measured by microarrays, are linear. A measurement thus consists of applying the chosen perturbations to each mRNA and measuring the responses of  $N$  concentrations of mRNA.

It takes more than one measurement to obtain solutions of the linearized kinetic equations, and suppose that  $M$  such measurements have been made so that the solutions yield the connectivity matrix. The positive matrix elements indicate activation of one mRNA by another, the negative elements yield repression of one mRNA by another, and the zero elements indicate no interactions. To obtain a unique solution the authors propose to choose among all possible solutions for the connectivity matrix by seeking the most sparsely connected network, the one with the smallest number of interactions among the mRNA. This yields a unique solution to the connectivity among the  $N$  species, but no justification is given for this particular choice. The method is tested on several model systems. They start with linear networks (see fig. 13.2) and consider 100, 200, 300, and 400 genes. For small numbers of measurements,  $M = 20$ , the errors made on finding connections between genes can be very large—in the thousands—but for each of these systems the errors are zero for  $M = 70$ . That is a large number of measurements; but the method as outlined works for small perturbations but not for large ones, for then the linearization is not applicable.

The concept of “sparse networks” is crucial to the success of this approach, but the authors were “unable to quantify this notion of sparseness and to pinpoint the critical size of a network as a function of the average number of connections” for which the method suffices. The method is not applicable for obtaining the connectivity of small, and likely highly connected, networks that govern specific biological functions.

An important consideration has been omitted in [3–5], which are devoted to this approach, and that is the usually rather large experimental errors associated with microarray measurements. It is important to know how such errors propagate in the calculations and their effects on the proper identification of the connectivity matrix. A simple example worked out in detail in section 12.5 shows the possible multiplicative effects of such errors in nonlinear kinetic equations. Until this problem is addressed, the approach of linearization must be viewed with caution.

### 13.3 Modeling of Reaction Mechanisms

We have described the modeling of reaction mechanisms for chemical systems in chapter 1. There we indicated the use of results of available experiments in regard to determining stoichiometries of elementary reactions and rate coefficients: the use of mass action kinetics, use of enzyme reaction mechanisms, guessing of missing reactions, and estimation (or even wild guessing) of missing rate coefficients. With this information

a reaction mechanism may be formulated, predictions from it may be compared with experiments, and new experiments may be suggested.

Some *tours de force* of these methods have been presented in several publications, (see [6,7] and references therein). The studies of Tyson and coworkers are focused on the kinetic analysis of the budding yeast cell cycle. The molecular mechanism of cell cycle control is known in more detail for budding yeast, *Saccharomyces cerevisiae*, than for any other eukaryotic organism. Many experiments have been done on this system over many years; there are about 125 references cited in [6]. The biological details are second to stressing the enormity of this task. The model has nearly twenty variables and that many kinetic equations, and there are about fifty parameters (rate coefficients, binding constants, thresholds, relative efficiencies). A fair number of assumptions need to be made in the cases of absence of any substantiating experimental evidence, and a fair number of approximations need to be made to simplify the kinetic equations. The complexity of this system is indicated in fig. 13.3 and its caption.

The authors make an extensive effort to compare with experiments, and there is substantial agreement (partly fitted) on variations in time during the cell cycle for several key species, both for the wild type and some mutants, and on the dependence of cell cycle time on growth rate and birth size, among others. For a large-scale modeling study of the microbial cell, see [7].

## 13.4 Boolean Networks

Suppose we simplify the expression of a gene as either “on” or “off.” In either state it may affect another gene. The “on”/“off” nature may be modeled by a Boolean function which has the value of 1 = “on” or 0 = “off.” In fig. 13.1, recall that the color red signifies increased expression and the color green signifies reduced expression. Hence we can think of fig. 13.1 being altered in this Boolean, binary simplification by substituting 1 for green and 0 for red (where such changes are scored as statistically significant).

Let us give a simple example of a Boolean network, taken from [8], shown in fig. 13.4. Part (a) of that figure shows the connectivity of the network; part (b) gives the logical (Boolean) rules; and part (c) the state transition table which results from the network structure and the logical rules. For example, consider the second line in the transition table, and we reason from the primed letters to the unprimed: If  $A' = 0$ , then  $B = 0$ ; if  $B' = 0$ , then  $A$  or  $C = 1$ ; if  $C' = 0$ , then  $A$  and  $B = 0$ . The task of reverse engineering is to find the network and rules that produce the observed transitions as displayed in fig. 13.4.

One way to proceed is to guess a Boolean network for the number of genes for which there are expression experiments, with input of known connectivities from prior experiments. Next, guesses need to be made of the logical rules, and here again the guesses may be educated by inspection of the experiments or given by random choice. Restrictions may be made, a priori, on the number of gene inputs on each gene. The calculations of a transition table for a given network and rules are fast due to the binary nature of the network. The calculated transition table can be compared visually with the experiments, which may give some hints on changes to be made in the network and the rules. However, a quantitative measure of agreement of the calculation with experiments is desirable, and such a measure is provided by information theory.

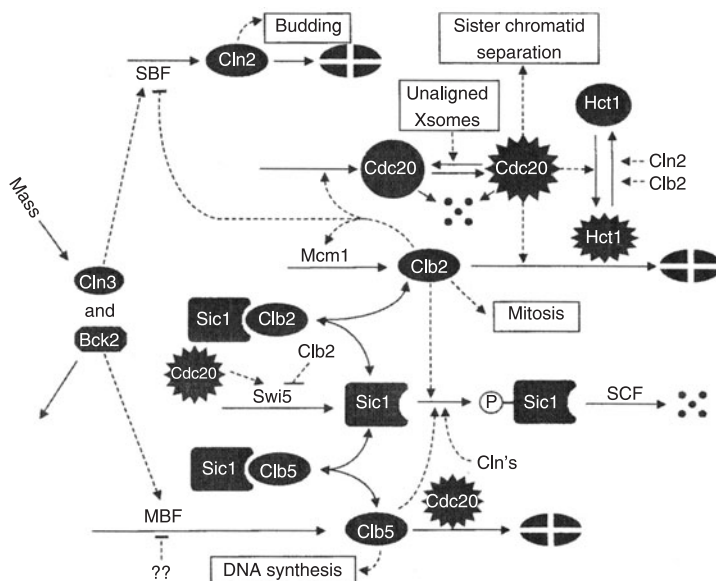


Fig. 13.3 Molecular model of the control of CDK (cyclin-dependent kinase) activities during the budding yeast cell cycle. We lump together redundant cyclins (Cln1 + Cln2 = “Cln2,” Clb1 + Clb2 = “Clb2,” and Clb5 + Clb6 = “Clb5”) and ignore Clb3 and Clb4. (Notice that we do not draw the kinase subunit, Cdc28, that is associated with each cyclin, because we assume it is in excess.) At the beginning of the cycle, the cell has few cyclin molecules, because the transcription factors (SBF, MBF, and Mcm1) that regulate cyclin synthesis are inactive. Clb-dependent kinases, in addition, are suppressed by a stoichiometric inhibitor (Sic1) and by efficient proteolysis of cyclin subunits. Cln3/Cdc28, which is present at low and nearly constant activity throughout the cycle, triggers a sequence of events leading ultimately to cell division. The sequence can be read from left to right. When the cell grows to a sufficiently large size, Cln3/Cdc28 and Bck2 activate SBF and MBF (by phosphorylation, so we assume), causing Cln2 and Clb5 to begin to accumulate. At first, Clb5 accumulates in inactive trimers, Clb5/Cdc28/Sic1, but Cln2/Cdc28 is not so inhibited. Rising Cln2/Cdc28 activity plays three important roles. First, it initiates bud formation. Second, it phosphorylates Sic1, priming the inhibitor for ubiquitination by SCF and ultimate degradation by the proteasome. Third, it inactivates Hct1, which, in conjunction with APC, was responsible for Clb2 degradation in G1 phase. When Sic1 is destroyed, Clb5/Cdc28 activity rises abruptly and drives the cell into S phase. These are the major physiological events associated with the Start transition. With Sic1 gone and Hct1 inactivated, Clb2-dependent kinase can begin to rise, with some lag, because Clb2/Cdc28 activates its own transcription factor, Mcm1. In addition, Clb2/Cdc28 inactivates SBF, so Cln2-dependent kinase activity begins to fall as Cln2 synthesis shuts off. At about the same time, MBF is inactivated, and the Clb5 level starts to fall. Rising Clb2/Cdc28 activity induces progress through mitosis. The metaphase–anaphase transition is regulated by a pair of proteins, Cdc20 and Hct1, that target substrates to the APC for ubiquitination. At metaphase they are inactive, but when DNA is fully replicated and all chromosomes are aligned on the metaphase plate, Cdc20 is activated. Indirectly Cdc20 promotes (1) dissociation of sister chromatids (anaphase A); (2) activation of Hct1, which conducts Clb2 to the APC, thereby initiating anaphase B and cell separation; and (3) activation of Swi5, the transcription factor for Sic1. With all CDK activity gone (except for a little associated with Cln3), Sic1 can make a comeback, and the cell returns to G1. (Figure and caption taken from [6], with permission.)

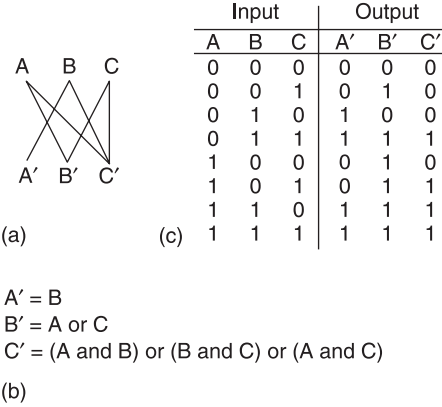


Fig. 13.4 A simple Boolean network. (a) Wiring diagram. (b) Logical Boolean rules. (c) State transition table which defines the network. The input column corresponds to time  $t$ , the output column to time  $t + 1$ . (Figure and caption taken from [8], with permission.)

We have already introduced that subject in chapter 9 (see eqs. (9.2)–(9.5)). There we chose a new measure of a correlation distance, the mutual information. A natural measure of such a distance between the probability distributions of two variables is the number of states jointly available to them (the size of the support set) compared to the number of states available to them individually. Thus, two probability distributions are close and the support set is small, if the knowledge of one predicts the most likely state of the other, even if there exists simultaneously a substantial number of other states. Hence from the gene expression distribution calculated in the transition table, and the measured expression distribution, a mutual information can be calculated and the search for an appropriate network proceeds until a minimum mutual information is obtained.

Other methods are available (see [9]). In that paper, calculated illustrations are given on the number of calculations necessary for Boolean network identification. For example, for a connectivity of 2 genes to each gene only 50 input–output pairs in the transition table had to be calculated from random guesses to identify a gene network of 320 genes.

### 13.5 Correlation Metric Construction for Genetic Networks

An application of time-lagged correlation analysis (CMC), following the work presented in chapter 7, to a gene network of 259 genes from a photosynthetic cyanobacterium has been reported by Schmitt et al. [10]. The gene network is perturbed with a flux of light and the responses of the genes, the gene transcription, is recorded on DNA microarrays at 20 min intervals for a period of 8–16 hours. An idealized example is shown in fig. 13.5; the presence of time lags indicates a cascade of biochemical reactions. The experimental setup and the two types of light intensity profiles used are shown in fig. 13.6. Each of the time measurements of the transcription responses is used to form time-lagged correlations with the time-varying light flux. Some results are shown in fig. 13.7. The first subdivision of the responses to the light flux are obtained from the maximum of the correlations; in group I that maximum occurred after a 20 min time lag, in group II after a 40 min time lag, and so on. Within each time



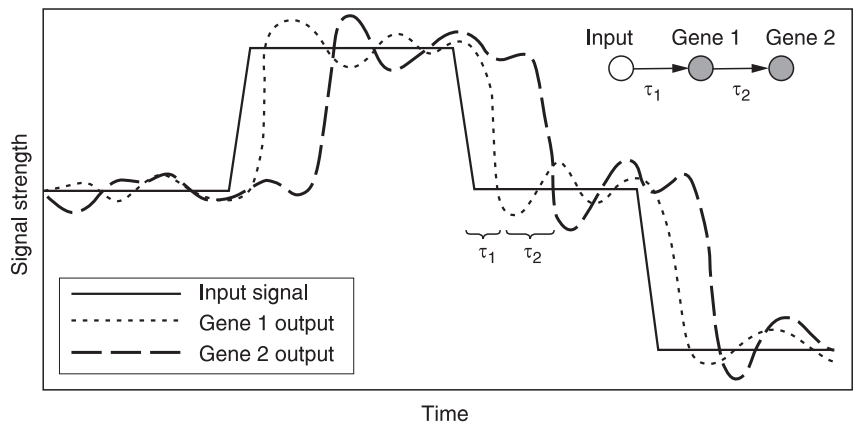


Fig. 13.5 Idealized gene expression of expected experimental results. The influx of light causes transcription in gene 1 with a time lag  $\tau_1$ . (Taken from [10], with permission.)

group, I–IV, further groupings are obtained by an analysis based on the profiles of the responses in the transcription. A listing of the genes in each group in fig. 13.7 is given in table 13.1.

For a summary of this work, we quote from [10]:

Although substantial effort is required to plan and perform this type of experiment, an enormous amount of information is obtained. The directionality of the resulting networks provides more information than clustering alone, and therefore allows the researcher to generate hypotheses based on the system structure. Additionally, it is important to consider

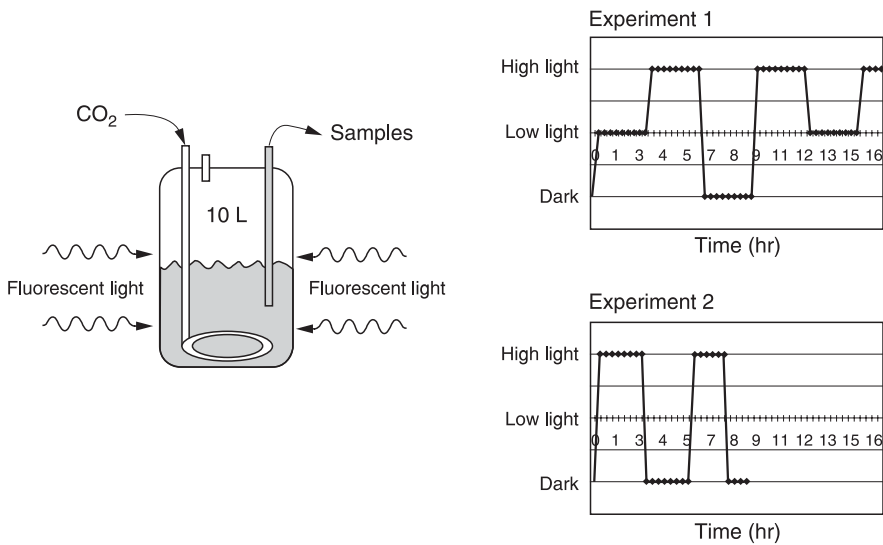


Fig. 13.6 Experimental setup and the two types of light intensity profiles used in the experiments. (Taken from [10], with permission.)

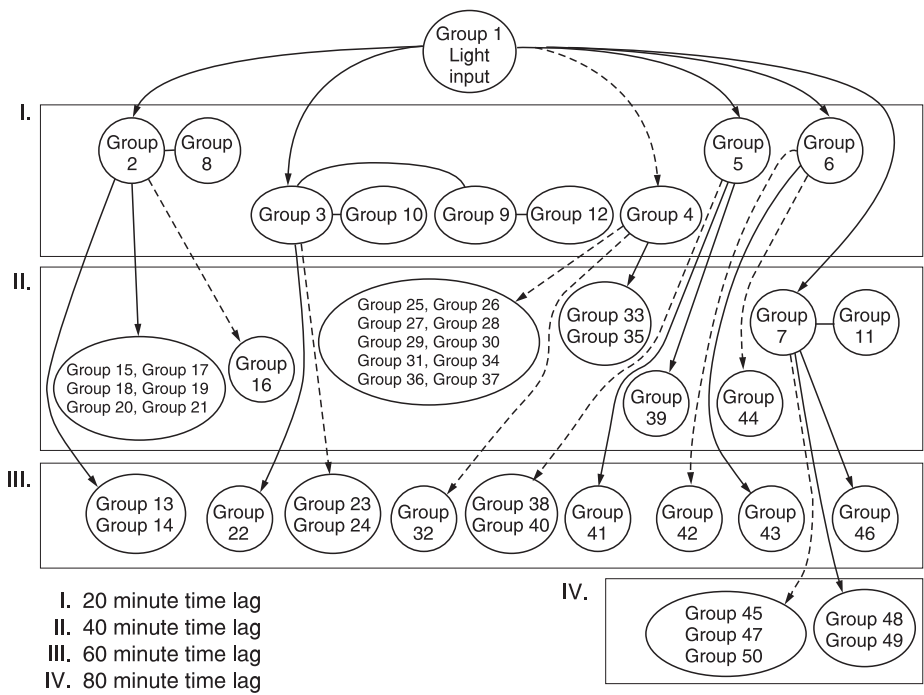


Fig. 13.7 Groupings of genes obtained from the time-lagged correlation analysis, groups I–IV, and further separations into numbered groups 2–50 obtained from an analysis of the profiles of the transcription responses. The genes in each numbered group are listed in table 13.1. Dark arrows indicate close correlation, a dotted line indicates inverse correlation, and a straight line indicates zero time lag. (Taken from [10], with permission.)

similarly expressed genes as potential regulon members. Regulons are sets of coregulated genes with common promoter regions differing from operons in that they are not necessarily sequentially oriented in the genome. To this end, genes with the same time-lagged correlations may be considered as good regulon candidates. . . . This suggests that dynamic studies of transcriptional behavior with significant number of time points can play a key role in understanding cellular regulation.

We cheerfully agree.  
See also [23] for another correlation analysis of a gene network.

13.6 Bayesian Networks

In chapter 7 we introduced concentration correlation functions of two species. In chapter 9 that concept was generalized to more than two species. Here we go all the way to the consideration of  $n$  species, with some limitations, by means of Bayesian networks [11–13]; we follow the presentation in [12] and [13], but much detail needs to be omitted. There is a very large literature on this subject and some of it is cited in [11–13] and in the review articles cited at the end of this chapter.

Table 13.1 Listing of the genes in each numbered group shown in fig. 13.7

<div>Group 2</div> <div>ycf59    slI1213    atpI atpC    atpA    atpL atpD    atpF atpG    atpH</div>			<div>Group 3</div> <div>slI0927    ycf48    psbE    psbF    psbL    psbJ    slI1070    slI1069 pphA    apcA    apcB    apcC    atpB    atpE    slr1331    slr1336 fpg    str0447    pgk    nbpl    chlP    sir0193    rpiA    ssl1263 ssr2998    acp    psaE    slI0822</div>						
<div>Group 4</div> <div>slr0967    slI1515    slI1009    ssl0832    slr0587 slr1232    ssr2078    slr1712    ssr2227    slr0582 slr0822    ssr0692    slr0581    slr0518    ssl1911</div>					<div>Group 5</div> <div>glnA</div>	<div>Group 6</div> <div>slI1835    guaA    slr1431    apcF hemE    gap2    argG</div>			
<div>Group 7</div> <div>cpcA    cpcB    cpcC2</div>			<div>Group 8</div> <div>rpl19    fus    ssl2245    etp    rpsla    accB    slI1234    ilvC slI1130    apcE    tufA    slr0752</div>						
<div>Group 9</div> <div>bioB    psbK    gyrB    gpx1    slr1020    slr1535    slr1160    slr1052    rps20    slr0981    rsp13    rpl136 clpP    purD    ndhH    hemB    slr1063    slr1618    slr1237    slr1062    SerA    slr1050    secF    slI1665 clpA    rbcL    nirA    murC    slr1176    slr1623    slr1619    slr1276    slr0483    slr1177    slr0982    slI1945 icd    rfbC    rbcX    petH    slr1277    slr1855    slr1624    slr1363    slr0773    slr1349    slr1051    slI1951 ppa    rps11    rfbE    rbcS    slr1617    slr2002    valS    slr1616    ycf3    trpE    slr1105    slr0776 psaD    secD    rpoC1    rfbF    slr1770    trpB    ycf23    thiC    rfbG    slr0484</div>									
<div>Group 10</div> <div>slI0843    slI0842</div>		<div>Group 11</div> <div>psaC    psaF    psaJ</div>			<div>Group 12</div> <div>cupB    slr2046    slr2047    slr2048    sigA ysf58    cpcG1    slr2052    dnaK    slr0654 gyIA    slI1921    pacS    slr2025    slr1464 crtQ-2    slr1462    fus</div>				
<div>Group 13</div> <div>psbM</div>	<div>Group 14</div> <div>slr1066    slr1070    slr1073    slr1071    slr1072</div>				<div>Group 15</div> <div>cpcD</div>		<div>Group 16</div> <div>slI1268</div>		
<div>Group 17</div> <div>ssr1480</div>	<div>Group 18</div> <div>ribF    ndbB</div>		<div>Group 19</div> <div>petG    ssI0294</div>		<div>Group 20</div> <div>petF</div>		<div>Group 21</div> <div>psbX</div>		
<div>Group 22</div> <div>slr1068</div>	<div>Group 23</div> <div>slr1396</div>	<div>Group 24</div> <div>slr0709</div>	<div>Group 25</div> <div>psaK    rps32    psal    psbl    psaL</div>						
<div>Group 26</div> <div>ssl2009</div>	<div>Group 27</div> <div>rps15</div>	<div>Group 28</div> <div>slI1837    slI1712    gap1    slr1821    slf1979    slr0821    ssl0242    rps21 slr0082    rps4    ndhj    slr1282    rpl21    slr0294    slr0038    slr0039 ssl2874    slr1406    ssr1528    ssl1046    slr0073    slr0765    slr1338    slI1343 rpl27</div>							
<div>Group 29</div> <div>serS    ccmK    natE</div>			<div>Group 30</div> <div>slr1721</div>	<div>Group 31</div> <div>sodB</div>	<div>Group 32</div> <div>slr0816    psbB    slr0907</div>				
<div>Group 33</div> <div>pppA    slI1386</div>		<div>Group 34</div> <div>ssr2130</div>	<div>Group 35</div> <div>slI0264    slI0135    hemD    slI0350    slI0232    slI0064    slI0103 slI0096    slI0136    slI0063    slI0062    slI0086    slI0084</div>						
<div>Group 36</div> <div>slr1926    ndhD1    slr0476</div>			<div>Group 37</div> <div>slr0784</div>	<div>Group 38</div> <div>slr2010</div>		<div>Group 39</div> <div>slr0886    glnB</div>		<div>Group 40</div> <div>slr1410</div>	
<div>Group 41</div> <div>slr1964</div>	<div>Group 42</div> <div>ssl1792</div>	<div>Group 43</div> <div>slr0915</div>	<div>Group 44</div> <div>slr0921</div>		<div>Group 45</div> <div>hisD    pxcA</div>			<div>Group 46</div> <div>hspA</div>	
<div>Group 47</div> <div>rpl24</div>	<div>Group 48</div> <div>slr2115</div>	<div>Group 49</div> <div>ycf46</div>	<div>Group 50</div> <div>ssl1045</div>						

Taken from [10], with permission.

A Bayesian network over a set of  $n$  species represents a joint probability distribution over all the species. The network is restricted to a directed acyclic graph; there are no feedforward or feedback loops, and the word “directed” dictates that the flow from one species to another is in one direction only. These are limitations to which we shall return later. A further specification is a conditional probability distribution for each variable given its parents in the graph  $G$ .

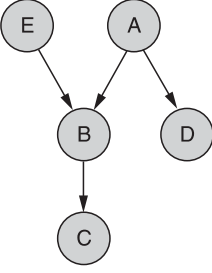


Fig. 13.8 Example of a Bayesian network consisting of five species. (Taken from [12], with permission.)

An example of a Bayesian network is given in fig. 13.8. The concentrations of the five species are random variables; each depends on the concentration of its parents, but not on other species. The parent of  $C$  is  $B$ , the parents of  $B$  are  $A$  and  $E$ . The connectivity of network structure in fig. 13.8 states that there is a probability  $P(A)$  that the concentration of  $A$  has a given value; a probability  $P(B|A,E)$  that  $B$  has a given value given the concentrations of  $A$  and  $E$ , the parents; and similarly for  $P(C|B)$ ,  $P(D|A)$ , and  $P(E)$ . The joint probability distribution, with the cited restrictions, is a product of these distributions, that is,

$$P(A, B, C, D, E) = P(A)P(B|A, E)P(C|B)P(D|A)P(E) \quad (13.1)$$

For the analysis of a gene network, gene expression is taken to be a probabilistic process and the level of gene expression of each gene is a random variable. The object of the analysis is the calculation of the joint probability distribution over the set of genes, not necessarily of all genes at a time but of a limited number, and to estimate its structure. There may be different but equivalent graphs that represent the observed distribution.

A generalization of eq. (13.1) is given by

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa_i^G) \quad (13.2)$$

where  $X_i$  is a set of random variables, and  $Pa_i^G$  is the set of parents of  $X_i$  in the graph  $G$ . A Bayesian graph implies certain properties of conditional independence (called Markov independence): each variable  $X_i$  is independent of its nondescendants, given its parents in  $G$ . Two Bayesian graphs may have the same set of dependences and independences. For example, in the graphs  $X \rightarrow Y$  and  $Y \rightarrow X$  the variables  $X$  and  $Y$  are dependent; the two graphs are not the same but they are equivalent. In comparing with experiments one cannot distinguish among equivalent graphs.

The comparison of an assumed Bayesian network with experiments proceeds by a learning process of the network; any prior knowledge of the system is entered deterministically, the other variables are random. To judge the progress of any learning, a statistically motivated scoring function needs to be introduced and evaluated. The search for a suitable network seeks an optimal score in this function. There are choices of doing all that; in [13] a score  $S(G : D)$  is defined as the posterior probability of a graph  $G$ , given the data (of the assumed network),  $P(G|D)$ :

$$S(G : D) = \log P(G|D) \quad (13.3)$$

The Bayesian probability formula is now invoked:

$$P(G|D) = \frac{P(D|G)P(G)}{P(D)} \tag{13.4}$$

which relates the posterior probability  $P(G|D)$  to the prior probability  $P(D|G)$ . We thus have

$$S(G : D) = \log P(D|G) + \log P(G) + C \tag{13.5}$$

where  $C$  is a constant independent of  $G$  and

$$P(D|G) = \int P(D|G, \Theta)P(\Theta)d\Theta \tag{13.6}$$

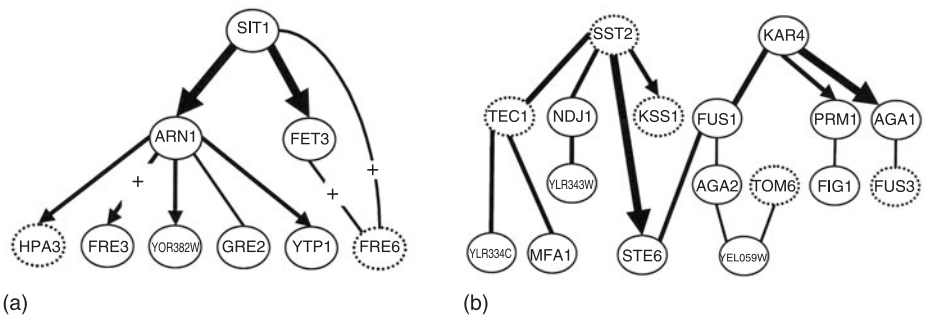


Fig. 13.9 Two subnetworks that show discovered features: (a) iron homeostasis; (b) mating response. The widths of the arcs correspond to the confidence in the feature. An edge is directed only if there is a high confidence in its orientation. Nodes circled with a dashed line correspond to genes that have been mutated in some of the samples. Arcs marked with a + sign are activators, with the size indicative of the confidence in this feature. (Figure and caption, with minor rewording, are taken from [13], with permission.)

Fig. 13.10 Global view and higher-order organization of modules. The graph depicts inferred modules (middle; number squares), their significantly enriched *cis*-regulatory motifs (right), and their associated regulators (left; ovals with black border for transcription factors or with green border for signal transduction molecules). Modules are connected to their significantly enriched motifs by solid blue lines. Module groups consisting of sets of modules that share a common motif, and their associated motifs, are enclosed in bold boxes. Only connected components that include two or more modules are shown. Motifs connected to all modules of their component are marked in bold. Modules are also connected to their predicted regulators. Red edges between a regulator and module are supported in the literature: either the module contains genes that are known targets of the regulator; or upstream regions of genes in the module are enriched for the *cis*-regulatory motif known to be bound by the regulator. Regulators tested experimentally are marked in yellow. Module groups are defined as sets of modules that share a single significant *cis*-regulatory motif. Module groups whose modules are functionally related are labeled (right). Modules belonging to the same module group seem to share regulators and motifs, with individual modules having different combinations of these regulatory elements. (Figure and caption, with minor changes, taken from [14], with permission.) (See color insert.)

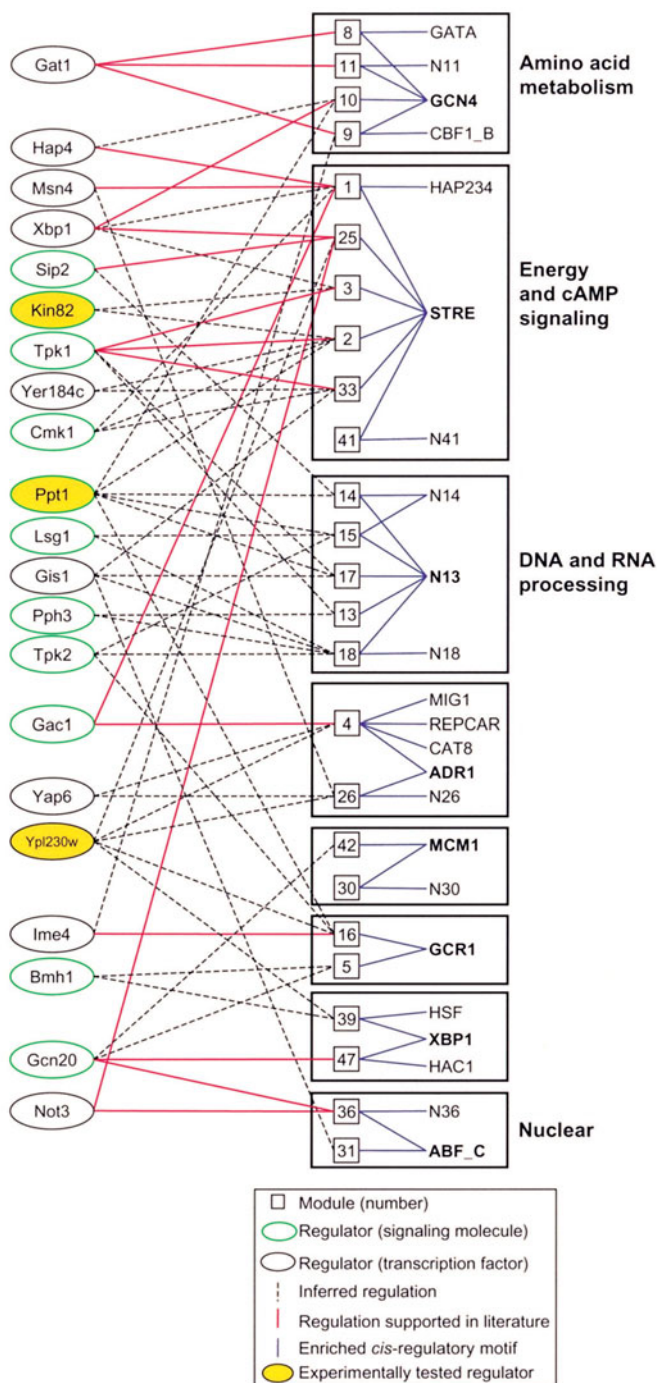


Fig. 13.10 Global view and higher-order organization of modules. See p. 218 for full caption.

is a marginal conditional probability, which averages the probability of the data over all possible parameters. Applied perturbations, such as mutations, deletions, and overexpression, require special considerations.

The calculations proceed as described, one stipulated Bayesian network after another, until an optimized score is attained. The optimized network, or the set of equivalent such networks, is then investigated for direct interactions of two, or more, species at a time to find so-called Markov neighbors; for separators, that is a species *Z* that separates species *X* and *Y*, and for activators and inhibitors. Such features are sought in all networks and a posterior probability of each feature, given the data, is established.

In [13] the authors chose microarray experiments on 565 genes from *Saccharomyces cerevisiae* to analyze by these methods. A great deal of useful information is obtained, of which we show two subnetworks inferred in some detail (fig. 13.9). The analysis can discover intracluster structure among correlated genes and intercluster connections between weakly correlated genes. The gene connections shown are the goal of such studies and are extremely useful for making predictions, for suggesting experiments, and ultimately for the treatment of genetic diseases. Are the subnetworks in fig. 13.9 sparse? That depends on an operational definition of sparse, yet to be formulated.

The Bayesian network approach is promising but there are limitations; we have mentioned already the restriction to acyclic graphs (no loops in the network) and the issue of Markovian independence. There are others: the presence of causal connectivity, as discussed in chapters 5 and 6, can be implied, can be indicated, but can not be determined. Finally, there is no rational basis, as yet, for connecting a Bayesian network with a chemical, biological, or genetic reaction mechanism: the equivalents of the concepts of temporal dynamics of reaction mechanisms, of rate coefficients, and of reversibility of elementary reactions are missing from Bayesian networks.

Following the unraveling of the connectivity of a genetic network, the next step is the experimental and theoretical study of the organization of the genome into regulatory modules [14]. A regulatory model is a given set of genes regulated together by a shared regulation program. It is assumed that the regulators are also transcriptionally regulated, and that their expression levels indicate their activity. The approach in this work is again based in part on Bayesian graphical methods, but goes further in ascertaining the presence and structure of modules; the computational requirements are substantial. We give here only an indication of the success of this approach by reproducing one of the figures from [14], fig. 13.10, with the sufficiently descriptive caption.

### 13.7 Some Other Illustrative Approaches

In [15] there is a “systematic determination of genetic architecture” based on a combination of clustering and statistical algorithms.

The responses to systematic perturbations of a metabolic network (galactose utilization) have been measured with microarrays and methods of quantitative proteomics. Nearly a thousand mRNA were identified in twenty systematic perturbations, which led to the identification of the physical interactions governing the cellular response to each perturbation [16].

An investigation of transcriptional regulatory networks in *Sacch. cerevisiae* is reported in [17]. The authors have determined the association of most of the transcriptional regulators with genes across the genome. They use statistical analysis of the measurements to show the presence of network motifs, and then demonstrate that an automated process can use these motifs to assemble a transcriptional regulatory network. For three review articles in this field, see [18–20].

We have encountered a number of difficult conceptual and practical problems in this chapter and throughout the entire book, and yet there are additional complications not treated. In a biological cell the concentrations of the various species may not be large, but many of the molecules (enzymes) may be large in size. Therefore molecular “crowding” [21] may occur; because of this effect, mass action kinetics does not apply and a form of fractal kinetics applies [22]. These problems have not yet been considered in the interpretation of most experimental investigations.

## References

- [1] Kohane, I. S.; Kho, A. T.; Butte, A. J. *Microarrays for an Integrative Genomics*; MIT Press: Cambridge, MA, 2003.
- [2] Eisen, M. B.; Spellman, P. T.; Brown, P. O.; Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 14863–14868.
- [3] Alter, O.; Brown, P. O.; Botstein, D. Singular value decomposition of genome-wide expression data processing and modelling. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10101–10106.
- [4] Yeung, M. K. S.; Tegner, J.; Collins, J. J. Reverse engineering gene networks using singular value decomposition and robust regression. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6163–6168.
- [5] Sontag, E.; Kiyatkin, A.; Kholodenko, B. N. Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data. *Bioinformatics* **2004**, *20*, 1877–1886.
- [6] Chen, K. C.; Csikasz-Nagy, A.; Gyorffy, B.; Val, J.; Novak, B.; Tyson, J. J. Kinetic analysis of a molecular model of the budding yeast cell. *Mol. Biol. Cell* **2000**, *11*, 369–391.
- [7] Ishii, N.; Robert, M.; Nakayama, Y.; Kanai, A.; Tomita, M. Toward large-scale modeling of the microbial cell for computer simulation. *J. Biotechnol.* **2004**, *113*, 281–294.
- [8] Liang, S.; Fuhrman, S.; Somogyi, R. REVEAL, a general reverse engineering algorithm for inference of genetic network architectures. *Pacific Symp. on Biocomputing* **1998**, *3*, 18–29.
- [9] Akutsu, T.; Miyano, S.; Kuhara, S. Identification of genetic networks from a small number of gene expression data under the Boolean network model. *Pacific Symp. on Biocomputing* **1999**, *4*, 17–28.
- [10] Schmitt, W. M. Jr.; Raab, R. M.; Stephanopoulos, G. *Genome Res.* **2004**, [www.genome-research.org/cgi/doi/10.1101/gr.2439804](http://www.genome-research.org/cgi/doi/10.1101/gr.2439804).
- [11] Jordan, M. I., Ed. *Learning in Graphical Models*; MIT Press: Cambridge, MA, 1999.
- [12] Friedman, N.; Lineal, M.; Nachman, I.; Pe’er, D. Using Bayesian networks to analyze expression data. *J. Comp. Biol.* **2000**, *7*, 601–620.
- [13] Pe’er, D.; Regev, A.; Elidan, G.; Friedman, N. Inferring subnetworks from perturbed expression profiles. *Bioinformatics* **2001**, *17*, (Suppl. 1), S215–S224.
- [14] Segal, E.; Shapira, M.; Regev, A.; Pe’er, D.; Botstein, D.; Koller, D.; Friedman, N. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* **2003**, *34*, 166–176.
- [15] Tavazole, S.; Hughes, J. D.; Campbell, M. J.; Cho, R. J.; Church, G. M. Systematic determination of genetic network architecture. *Nat. Genet.* **1999**, *22*, 281–285.
- [16] Ideker, T.; Thorsson, V.; Renish, J. A.; Christmas, R.; Buhler, J.; Eng, J. K.; Bumgarner, R.; Goodlet, D. R.; Aebersold, R.; Hood, L. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **2001**, *209*, 929–934.



- [17] Lee, T. I.; Rinaldi, N. J.; Roberts, F.; Odom, D. T.; Bar Joseph, Z.; Gerber, G. K.; Thompson, C. M.; Simon, I.; Zeilinger, J.; Jennings, E. G.; Murray, H. L.; Gordon, D. B.; Ren, B.; Wyrick, J. J.; Tagne, J.; Volkert, T. L.; Fraenkel, E.; Gofford, D. K.; Young, R. A. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **2002**, 298, 799–804.
- [18] Xia, Y.; Yu, H.; Jansen, R.; Seringhaus, M. Baxter, S.; Greenbaum, D.; Zhao, H.; Gerstein, M. Analyzing cellular biochemistry in terms of molecular networks. *Annu. Rev. Biochem.* **2004**, 73, 1051–1087.
- [19] Bolouri, H.; Davidson, E. H. Modeling transcriptional regulatory networks. *BioEssays* **2002**, 24, 1118–1129.
- [20] Eungdamrong, N. J.; Iyengar, R. Modeling cell signaling networks. *Biol. Cell* **2004**, 96, 355–362.
- [21] Schnell, S.; Turner, T. E. Reaction kinetics in intracellular environments with macro-molecular crowding: simulation and rate laws. *Progr. Biophys. Mol. Biol.* **2004**, 85, 235–260.
- [22] Kopelman, R. Fractal reaction kinetics. *Science* **1988**, 241, 1620–1626.
- [23] Remondini, D.; O'Connell, B.; Intrator, N.; Sedivy, J. M; Neretti, N.; Castellani, G. C.; Cooper, L. N. Targeting c-Myc-activated genes with a correlation method: detection of global changes in large gene expression network dynamics. *Proc. Natl. Acad. Sci. USA* **2005**, 102, 6902–6906.

# Index

---

Page numbers in *italic* refer to figures; page numbers in **boldface** signify entries in tables.

- activated complex theory, 8
- activity coefficients, 8
- AND gate, 37, 43
- Arrhenius equation, 7–8
- autocatalysis, 125
- autocatalytic species, 132
- autocorrelation function, 25
  
- Belousov–Zhabotinsky (BZ) reaction, 125–127
- bifurcation analysis, 129–131, *129*, *130*
- bifurcation diagrams, *115*, 143–147, *145*, *146*
- binary coding, 104–105
- biochemical reaction network (BRN), 40
- bioinformatics, 207–209
  - Bayesian networks, 215–220, *216*, *217*, *218*, *219*
  - Boolean networks, 211–213, *213*
  - clustering, 208, 209
  - correlation metric construction for genetic networks, 213–215, *214*, *215*
  - linearization, 209–210, *210*
  - modeling reaction mechanisms, 210–211, *212*
  - other approaches, 220–221
  - bistable reaction systems and bistable electronic switches, 37–38
  - blue fluorescent protein (BFP), 23
  
- capillary electrophoresis (CE), 13–15
- capillary isoelectric focusing (CIEF), 14
- capillary zone electrophoresis (CZE), 14
- catalyst, definition, 3
- causal connectivity, definition, 3
- cellular dynamics, 12
- chemical neurons, 34–37, *36*
- chlorite–iodide reaction, 129–131, *129*, *130*, 160
  - bifurcation diagrams, 163–164, *163*
  - concentration shift destabilization, 161–162, *162*, **162**
  - concentration shift regulation, 161, **161**
  - model, 164–165, *165*
  - phase shift of oscillations, 164
  - qualitative pulsed species response, 160–161, **161**
- classical identification, 4–6
- complex systems, deducing reaction mechanisms, 2

- computations by macroscopic kinetic systems, 34
- chemical neurons and logic gates, 34–37, 36
- computational functions, 40–44, 41
- implementation of computers, 37–40, 38, 39
- concentration
  - correlation function method, 2
  - relative change, 47
  - stationary state, 47
- continuous-flow stirred tank reactor (CSTR), 37–39, 38, 39, 59
- control, definition, 3
- controllability, 5
- Coomassie Blue stain, 21
- correlation metric construction (CMC)
  - experimental test, 87–96, 88, 90
  - matrix summary of distance matrix, 92
  - MDS diagram, 93
  - time-lagged correlation function, 91
- genetic networks, 213–215, 214, 215
- theory of statistical construction of reaction mechanisms, 65–85, 68, 72, 77, 78, 81
- cross-correlation function, 25
- cross-shaped diagram, 131, 144
- cyan fluorescent protein (CFP), 23
- cycle species, 132
- cyclic pathways, 135–136
- cyclin-dependent kinase (CDK), 212
- density estimation
  - entropy metric construction (EMC), 97–102, 100, 101
- deterministic kinetics, 7–8
- disappearance processes, 193
- disordered kinetics, 9
- DNA microarrays, 18–20, 19
- Drazer–Zanette experiment, 176–180, 181
- Easterby's equation, 187
- electro-osmotic flow (EOF), 13–14
- electrospray ionization (ESI), 16
- elementary reaction, definition, 3
- entropy, 98
- entropy metric construction (EMC), 97–102, 100, 101
- entropy reduction method (ERM), 102
- enzyme binding and rate coefficients (EBR), 114–115
- exit species, 132
- feedback mechanisms, 125
- feedback species, 133
- fluctuations, 9
- fluorescence localization after photobleaching (FLAP), 24
- fluorescence loss in photobleaching (FLIP), 24
- fluorescence recovery after photobleaching (FRAP), 24
- fluorescent dyes, 21
- fluorescent imaging, 23
  - fluorescence correlation spectroscopy (FCS), 25–26, 25
  - fluorescence resonance energy transfer (FRET), 26–27
  - fluorescence speckle microscopy (FSM), 24
- future trends, 27
- photobleaching and photoactivation, 23–24
- total internal reflection fluorescence microscopy (TIRFM), 25
- Fourier Transform Ion Cyclotron Resonance (FT-ICR) MS, 17
- fructose 6-phosphate/fructose 1,6-biphosphate cycle, 40–43, 42
- gas chromatography-mass spectrometry (GC-MS), 13
- genetic algorithm method, 2
- genetic algorithms (GAs), 104–105, 122
  - basic techniques
    - crossover, 105
    - elitism, 105
    - mutation, 105
    - roulette wheel selection, 105
    - termination, 106
- evolutionary development of biochemical oscillatory reaction mechanisms, 112–118, 117, 118
  - enzyme binding constants, 114
- flux regulation selection, 106–112, 106
  - limiting network diagrams, 111
  - noncompetitive binding, 107
- systematic determination of reaction mechanism and rate coefficients, 119–122
- genome, 207
- gluconeogenesis, 43
- glycolysis, 4, 42–43, 113
  - experimental test of correlation metric construction (CMC), 87–96, 88, 90
  - matrix summary of distance matrix, 92
  - MDS diagram, 93
  - time-lagged correlation function, 91

- experimental test of pulse perturbation
    - method, 58–64, 59, 60, 61, 63
  - pathway, 53–56, 54, 55, 56
- green fluorescent protein (GFP), 23
- Green functions, 185, 193
- high-performance anion-exchange
  - chromatography (HPAEC), 12, 13
- high-pressure liquid chromatography (HPLC), 12–13
  - proteome analysis, 21
- Hopf bifurcation, 129, 130, 131, **143**
- horseshoe effect, 83
- hydrophilic-interaction chromatography (HILIC), 13
- input negative feedback, 137
- inverse self-regulation, 142
- iodate–arsenous acid reaction, 37, 38, 39
- ion trap mass analysers, 17
- ionic strength, 8
- isobaric tags, 22
- isotope-coded affinity tags (ICAT), 22
- Jacobian matrix elements (JMEs), 128–129
- Kalman filtering, 5
- kinetic processes
  - disordered kinetics, 9
  - fluctuations, 9
  - macroscopic, deterministic kinetics, 7–8
- Langmuir adsorption isotherms, 176
- lifetime distributions of species, 171–173, 205
  - response experiments, 173–182
  - errors, 192–197, 194, 195
- LISREL, 5
- logic gates, 34–37, 36, 40–43, 42
- marked species, 198
- mass action law, 46
- mass action rate expressions, 7
- mass spectrometry (MS)
  - principles and applications, 15–16
  - detectors, 17–18
  - HPLC detectors, 13
  - mass analyzers, 16–17
- mass-to-charge ratio ( $m/z$ ) of ions, 16
- matrix-assisted laser desorption/ionization (MALDI), 16
- metabolome, 12
- methodologies, 11, 27
- fluorescent imaging, 23–27
- genome-wide analyses of mRNA and proteins, 18
  - DNA microarrays, 18–20, 19
  - proteome analysis, 20–23, 22
- mass spectrometry (MS), 15–16
  - detectors, 17–18
  - mass analyzers, 16–17
- metabolite concentration measurements, 11–12
  - capillary electrophoresis (CE), 13–15, 15
  - high-pressure liquid chromatography (HPLC), 12–13
- metric, requirements, 98
- micellar electrokinetic capillary chromatography (MECC), 14
- molality, 8
- mRNA, 207
  - linearization, 209–210, 210
- multidimensional scaling (MDS), 73–82, **75**, 76, 78, 81
  - glycolysis, 93
- multivariate systems, 5–6
- NAND gate, 66, 78
- network performance evaluation, 109–110
- neural networks, 40
- NFT mechanism, 119–120
- nonlinear systems, 5
- NOR gate, 42
- normal self-regulation, 142
- normalized correlation functions, 67
- Nyquist critical frequency, 6
- observability, 5
- OR gate, 43
- oscillatory reactions, 2, 125–127
  - concepts and theoretical constructs, 127–128
  - bifurcation analysis, 129–131, 129, 130
  - categorization of oscillatory reaction, 136–137
  - Jacobian matrix elements (JMEs), 128–129
  - role of species, 131–133
  - stoichiometric network analysis (SNA), 133–136, 134, 136
- deduction of mechanism examples, 151
  - chlorite–iodide reaction, 160–165
  - peroxidase–oxidase (PO) reaction, 151–160, **152**

oscillatory reactions (*continued*)

- experiments providing mechanism
    - information, 138
    - amplitude relations, 138, 140
  - bifurcation diagrams, 143–147, 145, 146
  - characterization of species, 138, 139
  - concentration shift destabilization, 143
  - concentration shift regulation, 141–142
  - delay experiments, 148
  - external periodic perturbation, 150
  - other methods, 151
  - phase relations, 139–141
  - phase response experiments, 149–150
  - pulsed species response, 147–148, **147**
  - quenching, 148–149
  - stabilization of unstable steady state, 150
  - limits of stoichiometric network analysis, 166
- output negative feedback, 137
- ozone, formation and decomposition by light in upper atmosphere, 2
- pair correlation function, 97
- pair probability distribution, 99
- parallel computing, 40
- pattern recognition, 39
- pentose cycle, 4
- peroxidase–oxidase (PO) reaction, 151–152, 156–157, **157**
- concentration shift destabilization, 154–155, 155
  - concentration shift regulation, 153–154, 154, **154**
- model, 157–160, 158, **159**
- phase shifts of oscillations, 152, **153**
- qualitative pulsed species response, 155, 155
- quench experiments, 156, 156
- relative amplitudes of oscillations, 152, 153
- phase response curve (PRC), 149
- phase transition curve (PTC), 149
- phenotype, 11
- phosphorylation/dephosphorylation cycle, 41
- photoactivation, 23–24
- photobleaching, 23–24
- Polanyi relations, 177
- polyacrylamide gel electrophoresis, 21
- primary salt effect, 8
- proteome analysis, 20–23, 22
- pulse method, 2

## pulse perturbations

- experimental test, 58–64, 59, 60, 61, 63
  - glycolytic pathway, 53–56, 54, 55, 56
  - theory, 46–53, 47, 48, 50, 51, 52
- quadrupole mass analyzers, 16
- quenching, 148–149
- rate coefficients, 8
- reachability, 5
- reaction–diffusion systems, 197–204
- reaction mechanisms, definition, 3
- reaction pathway, definition, 3
- reactions, 1
- recovery species, 133
- red fluorescent protein (RFP), 23
- relative change in concentration, 47
- replication processes, 193
- reverse-phase (RP) HPLC, 13
- RNA, 207. *See also* mRNA
- saddle-node bifurcations, 130
- selected ion monitoring (SIM), 16
- SEPATH, 5
- sign-symbolic shift matrix, 139, **140, 141, 142**
- silver staining, 21
- single nucleotide polymorphisms (SNP), 18
- stationary state concentrations, 47
- statistical construction of reaction mechanisms, 65–85, 68, **72**, 77, 78, 81
- stochastic elements in reactions, 5
- stoichiometric equations, 1
- stoichiometric network analysis (SNA), 131, 133–136, 134, 136
- limits, 166
- susceptibility functions, 185
- time-lagged correlation function, 67, 69, 70, 73
- glycolysis, 91
- time-of-flight (TOF) mass analyzers, 16–17
- total internal reflection fluorescence microscopy (TIRFM), 25
- tracers, 170
- transformation processes, 193
- transit time distributions, 182–188
- extracting kinetic information from experimental data, 188–192
- Turing machines, 40
- yellow fluorescent protein (YFP), 23
- Zeldovich–Roginskii model, 178